PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12Q 1/68

A2

(11) International Publication Number: **WO 00/47771**

(43) International Publication Date: 17 August 2000 (17.08.00)

(21) International Application Number: PCT/IB00/00184

(22) International Filing Date: 11 February 2000 (11.02.00)

(30) Priority Data:

60/119,917	12 February 1999 (12.02.99)	US
09/275,267	23 March 1999 (23.03.99)	US
60/133,200	7 May 1999 (07.05.99)	US

(71) Applicant *(for all designated States except US)*: GENSET [FR/FR]; Intellectual Property Dept., 24, rue Royale, F–75008 Paris (FR).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: BLUMENFELD, Marta [FR/FR]; 5, rue Tagore, F–75013 Paris (FR). BOUGUELERET, Lydie [FR/FR]; 108, avenue Victor Hugo, F–92170 Vanves (FR). CHUMAKOV, Ilya [FR/FR]; 196, rue des Chevrefeuilles, F–77000 Vaux–le–Penil (FR).

(74) Common Representative: GENSET; Intellectual Property Dept., 24, rue Royale, F–75008 Paris (FR).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING GENES INVOLVED IN ARACHIDONIC ACID METABOLISM

(57) Abstract

The invention provides polynucleotides including biallelic markers derived from genes involved in arachidonic acid metabolism and from genomic regions flanking those genes. Primers hybridizing to regions flanking these biallelic markers are also provided. This invention also provides polynucleotides and methods suitable for genotyping a nucleic acid containing sample for one or more biallelic markers of the invention. Further, the invention provides methods to detect a statistical correlation between a biallelic marker allele and a phenotype and/or between a biallelic marker haplotype and a phenotype.

This Page Blank (uspto)

BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING GENES INVOLVED IN ARACHIDONIC ACID METABOLISM

FIELD OF THE INVENTION

The present invention is in the field of pharmacogenomics, and is primarily directed to biallelic markers that are located in or in the vicinity of genes, which have an impact on arachidonic acid metabolism and the uses of these markers. The present invention encompasses methods of establishing associations between these markers and diseases involving arachidonic acid metabolism such as inflammatory diseases as well as associations between these markers and treatment response to drugs acting on arachidonic acid metabolism. The present invention also provides means to determine the genetic predisposition of individuals to such diseases and means to predict responses to such drugs.

BACKGROUND OF THE INVENTION

The metabolites of arachidonic acid and related fatty acids, collectively termed eicosanoids, exhibit a wide range of biological activities affecting virtually every organ system in mammals. Eicosanoids are among the most important chemical mediators and modulators of the inflammatory reaction and contribute to a number of physiological and pathological processes (See Hardman J.G., Goodman, Gilman A., Limbird L.E.; *Goodman & Gilman's The Pharmacological Basis of Therapeutics*, 9th edition, McGraw-Hill, N.Y., 1996).

Physiology, Pathophysiology and Pharmacological Importance of the Eicosanoids

The eicosanoids are extremely prevalent and have been detected in almost every tissue and body fluid. These lipids contribute to a number of physiological and pathological processes including inflammation, smooth muscle tone, hemostasis, thrombosis, parturition and gastrointestinal secretion. Once synthesized in response to a stimulus, the eicosanoids are not stored to any significant extent but are released immediately and act locally. After they act, they are quickly metabolized by local enzymes to inactive forms. Accordingly, the eicosanoids are categorized as autocrine agents or local hormones. They alter the activities of the cells in which they are synthesized and of adjoining cells. The nature of these effects may vary from one type of cell to another, in contrast with the more uniform actions of global hormones such as insulin, for example. Therefore, the eicosanoids, as local chemical messengers, exert a wide variety of effects in virtually every tissue and organ system.

The principal eicosanoids are the prostaglandins (PG), the thromboxanes (TX) and the leukotrienes (LT), though other derivatives of arachidonate, for example lipoxins, are also produced.

They fall into different classes designated by letters and the main classes are further subdivided and designated by numbers.

Inflammatory and immune responses

Eicosanoids are lipid mediators of inflammation and play a central, often synergistic, role in numerous aspects of inflammatory responses and host defense. Prostaglandins and leukotrienes are released by a host of mechanical, thermal, chemical, bacterial, and other insults, and they contribute importantly to the genesis of the signs and symptoms of inflammation. The ability to mount an inflammatory response is essential for survival in the face of environmental pathogens and injury, although in some situations and diseases the inflammatory response may be exaggerated and sustained for no apparent beneficial reason. This is the case in numerous chronic inflammatory diseases and allergic inflammation. Acute allergic inflammation is characterized by increased blood flow, extravasation of plasma and recruitment of leukocytes. These events are triggered by locally released inflammatory mediators including eicosanoids and more particularly leukotrienes. The leukotrienes generally have powerful effects on vascular permeability and the leukotriene LTB_4 is a potent chemoattractant for leukocytes and promotes exudation of plasma. The prostaglandins PGE_2 and PGI_2 markedly enhance edema formation and leukocyte infiltration in the inflamed region. Moreover, they potentiate the pain-producing activity of bradykinin.

The participation of arachidonic acid (AA) metabolism in inflammatory diseases such as rheumatoid arthritis, asthma and acute allergy is well established. Prostaglandins have been involved in inflammation, pain and fever. Pathological actions of leukotrienes are best understood in terms of their roles in immediate hypersensitivity and asthma. Lipoxygenases, e.g., 5-lipoxygenase (5-LO), 12-lipoxygenase (12-LO), 15-lipoxygenase A (15-LOA), and 15-lipoxygenase B (15-LOB), have been implicated in the pathogenesis of a variety of inflammatory conditions such as psoriasis and arthritis.

Cardiovascular system

The prostaglandins PGEs, PGF_2 and PGD_2 cause both vasodilation and vasoconstriction. Responses vary with concentration and vascular bed. Systemic blood pressure generally falls in response PGEs, and blood flow to most organs, including the heart, is increased. These effects are particularly striking in some hypertensive patients. Cardiac output is generally increased by prostaglandins of the E and F series. The importance of these vascular actions is emphasized by the participation of PGI_2 and PGE_2 in the hypotension associated with septic shock. The prostaglandins also have been implicated in the maintenance of patency of the ductus arteriosus. Thromboxane synthase (TXA2), also known as CYP5, is a potent vasoconstrictor. Leukotriene C_4 synthase (LTC_4)

and the leukotriene LTD_4 cause hypotension. The leukotrienes have prominent effects on the microvasculature. $LTC_4$and LTD4 appear to act on the endothelial lining of postcapillary venules to cause exudation of plasma; they are more potent than histamine in this regard. In higher concentrations, LTC4 and LTD4 constrict arterioles and reduce exudation of plasma.

Blood/Platelets

Prostanoids including prostaglandins and thromboxanes exhibit a wide variety of actions in various cells and tissues to maintain local homeostasis in the body. Eicosanoids modify the function of the formed elements of the blood. PGI2 controls the aggregation of platelets in vivo and contributes to the antithrombogenic properties of the intact vascular wall.

TXA2 is a major product of arachidonate metabolism in platelets and, as a powerful inducer of platelet aggregation and the platelet release reaction, is a physiological mediator of platelet aggregation. Pathways of platelet aggregation that are dependent on the generation of TXA2 are sensitive to the inhibitory action of aspirin, which inhibits the cyclooxygenase (COX) pathway. There has been considerable interest in the elucidation of the role played by prostaglandins and TXA2 in platelet aggregation and thrombosis and by PGI_2 in the prevention of these events. The platelet thromboxane pathway is activated markedly in acute coronary artery syndromes and aspirin is beneficial in the secondary prevention of coronary and cerebrovascular diseases. PGI that is generated in the vessel wall may be the physiological antagonist of this system; it inhibits platelet aggregation and contributes to the nonthrombogenic properties of the endothelium. According to this concept, PGI_2 and TXA2 represent biologically opposite poles of a mechanism for regulating platelet-vessel wall interaction and the formation of hemostatic plugs and intraarterial thrombi. There is interest in drugs which inhibit thromboxane synthase and modulate PGI2 production.

Smooth muscle

Prostaglandins contract or relax many smooth muscles beside those of the vasculature. The leukotrienes contract most smooth muscles. In general, PGFs and PGD2 contract and PGEs relax bronchial and tracheal muscle. LTC4 and LTD4 are bronchoconstrictors. They act principally on smooth muscle in peripheral airways and are 1000 times more potent than histamine both *in vitro* and *in vivo*. They also stimulate bronchial mucus secretion and cause mucosal edema. A complex mixture of chemical messengers is released when sensitized lung tissue is challenged by the appropriate antigen. Various prostaglandins and leukotrienes are prominent components of this mixture. Response to the leukotrienes probably dominates during allergic constriction of the airway. Evidence for this conclusion is the ineffectiveness of inhibitors of cycloxygenase and of histaminergic antagonists in the

treatment of human asthma and the protection afforded by leukotriene antagonists in antigen induced bronchoconstriction. A particularly important role for the cysteinyl-leukotrienes (LTC4, LTD4, and LTE4) has been suggested in pathogenesis of asthma, which is now recognized as a chronic inflammatory condition. They are potent spasmogens causing a contraction of bronchiolar muscle and an increase in mucus secretion.

Gastric and intestinal secretions

PGEs and PGI2 inhibit gastric acid secretion stimulated by feeding, histamine or gastrin. Mucus secretion in the stomach and small intestine is increased by PGEs. These effects help to maintain the integrity of the gastric mucosa and are referred to as the cytoprotectant properties of PGEs. Furthermore, PGEs and their analogs inhibit gastric damage caused by a variety of ulcerogenic agents and promote healing of duodenal and gastric ulcers. Cytoprotection is of therapeutic importance and PGE_1 analogs are used for the prevention of gastric ulcers.

Kidney and urine formation

Prostaglandins modulate renal blood flow and may serve to regulate urine formation by both renovascular and tubular effects. Increased biosynthesis of prostaglandins has been associated with Bartter's syndrome, a rare disease, characterized by urinary wasting of K^+.
Leukotrienes have been involved in the pathophysiology of glomerular immune injury.

Reproduction and parturition

Much interest is attached to the possible involvement of prostaglandins in reproductive physiology. Lowered concentrations of prostaglandins in semen have been implicated in male infertility. Prostaglandins are also thought to contribute to the symptoms of primary dysmenorrhea. Inhibitors of cyclooxygenase are effective in relieving the symptoms of this condition. Elevated levels of prostaglandins are involved in onset of labor. Inhibitors of cyclooxygenase increase the length of gestation and interrupt premature labor.

Cancer metastasis

Tumors in animals and certain spontaneous human tumors are accompanied by increased concentrations of local or circulating prostaglandins. Eicosanoids have been shown to be involved in various aspects of neoplasia including cell transformation, tumor promotion, tumor cell growth, and metastasis. Some studies have implicated platelet aggregation and the effects of prostaglandins and hydroxyeicosatetraenoic acid (12-HETE) in the hematogenous metastasis of tumors.

Many of the products of arachidonic acid metabolism are potent mediators of physiological responses and contribute to disorders of development, cellular function, tissue repair, and host defenses in a number of diseases.

Arachidonic Acid Metabolism And Biosynthesis Of Eicosanoids

The primary source of eicosanoids in mammalian systems is the metabolic products of arachidonic acid. After stimulation by trauma, infection, or inflammation, translocated phospholipases, especially phospholipase A_2, act on membrane phospholipids to liberate arachidonic acid. Once released, arachidonate is metabolized to oxygenated products by several distinct enzyme pathways, including cyclooxygenases, several lipoxygenases, and cytochrome P450s (CYP). The specific enzyme pathway involved determines, which products are formed. See Figure 1.

Release of arachidonic acid from cell membranes and its regulation

The eicosanoids are a family of substances produced from the polyunsaturated fatty acid arachidonic acid, which is present in plasma-membrane phospholipids. The first rate-limiting step in the biosynthesis of eicosanoids is the release of arachidonic acid from the membrane, a process that is mainly catalyzed by cytosolic phosholipase A_2 ($cPLA_2$). The synthesis of eicosanoids begins when a stimulus such as a hormone, a neurotransmitter, a drug or a toxic agent activates cytosolic phospholipase A_2. This arachidonic acid specific phospholipase plays a major role in the cell signaling events that initiate the arachidonate cascade. One important trigger of arachidonate release and eicosanoid synthesis involves tissue injury and inflammation.

The activities of many enzymes are regulated by calmodulins (CAL) that serve as calcium sensors in eukaryotic cells. The binding of Ca^{2+} to multiple sites in calmodulin induces a major conformational change that converts it from an inactive to an active form. Activated calmodulin then binds to many enzymes and target proteins in the cell, modifying their activities and thereby regulating various metabolic pathways. Calmodulins are involved in a number of processes regulated by Ca^{2+} including smooth muscle contraction, neurotransmission, apoptosis, cell cycle progression and gene expression. Calmodulins also participate in the regulation of arachidonate release. They directly stimulate cytosolic phospholipase A_2, whereas calmodulin antagonists inhibit enzyme activity and the release of arachidonic acid.

Annexins (ANX) are a family of multifunctional calcium and phospholipid-binding proteins, they belong to a family of proteins that interact with phospholipids in a Ca^{2+} dependant manner. Annexins have been implicated in the pathogenesis of benign and malignant neoplasms of different origins. Moreover, several annexins have also been involved in autoimmune diseases such as systemic

lupus erythematosus, rheumatoid arthritis and inflammatory bowl disease. Numerous physiological functions have been attributed to annexins including regulation of membrane traffic during exocytosis and endocytosis, mediation of cytoskeletal-membrane interactions, membrane receptor function, regulation of membrane-dependent enzymes, mitogenic signal transduction, transmembrane ion channel activity, cell-cell adhesion, antiinflammatory properties, inhibition of blood coagulation and inhibition of phospholipase A_2. Annexins have been suggested as regulators of prostaglandin metabolism and of the arachidonate cascade as a result of their inhibitory effect on phospholipase A_2. It is still a matter of debate as to whether inhibition of phospholipase A2 is the result of calcium-dependent sequestration of phospholipids (substrate depletion mechanism) or a direct effect of the annexins acting via protein-protein interactions. Calpactin I (light chain) is the cellular ligand of annexin II and induces its dimerization. Annexin II and calpactin I (CALPA) constitute a calcium binding complex composed of two light chains (calpactin I) and two heavy chains (annexin II). Calpactin I may function as regulator of annexin II phosphorylation.

The activities of phospholipase A_2, annexins and calmodulins are common points of regulation in the formation of all eicosanoids.

Downstream of phospholipase A_2, the varying eicosanoid-pathway enzymes found in particular cell types determine which eicosanoids are synthesized in response to particular stimuli.

Cyclooxygenase pathway

This pathway initiated by cyclooxygenase (COX) leads ultimately to formation of the cyclic endoperoxides, prostaglandins (PG), and thromboxanes (TX).

There are two isoforms of the cyclooxygenase, COX-1 and COX-2. The former is constitutively expressed in most cells. In contrast, COX-2 is not normally present but may be induced by certain factors such as cytokines and growth factors. The cyclooxygenases have two distinct activities: an endoperoxidase synthase activity that oxygenates and cyclizes the unesterified precursor fatty acid to form the cyclic endoperoxide PGG and a peroxidase activity that converts PGG to PGH. PGG and PGH are chemically unstable, but they can be transformed enzymatically into a variety of products, including PGI, TXA2, PGE, PGF or PGD. Isomerases lead to the synthesis of PGE_2 and PGD_2, whereas PGI_2 is formed from PGH_2 through prostacyclin synthase. TXA2 is formed by thromboxane synthase. Although most tissues are able to synthesize the PGG and PGH intermediates from free arachidonate, the fate of these precursors varies in each tissue and depends on the complement of enzymes that are present and on their relative abundance. For example, lung and spleen are able to synthesize the whole range of products. In contrast, platelets contain thromboxane synthase as the principal enzyme that metabolizes PGH, while endothelial cells contain primarily prostacyclin synthase.

Lipoxygenase pathways

Lipoxygenases are a family of cytosolic enzymes that catalyze the oxygenation of fatty acids to corresponding lipid hydroperoxides. See Figure 1. Arachidonate is metabolized to HPETE (hydroperoxyeicosatetraenoic acid), which is then converted either enzymatically or non-enzymatically to 12-HETE (hydroxyeicosatetraenoic acid). HPETEs may further be converted to hepoxilins and lipoxins. Lipoxygenases differ in their specificity for placing the hydroperoxy group, and tissues differ in the lipoxygenases they contain. These enzymes are referred to as 12-, 15-, 5- and 8-lipoxygenases according to the oxygenation sites in arachidonic acid as substrate.

The lipoxygenases catalyze reactions and generate products of potential relevance to membrane remodeling, cell differentiation and inflammation. Products of the 15-LO pathway could contribute to the pathophysiology of allergic airway inflammation while products of the 12-LO pathway have been implicated in cancer metastasis, psoriasis and inflammation.

Various biological activities have been reported for the 12-lipoxygenase metabolites of arachidonic acid. As other eicosanoids, they are important chemical mediators and modulators of the inflammatory reaction. 12-HETE is the major arachidonic acid metabolite of 12-lipoxygenase and seems to be implicated in a wide-spectrum of biological activities such as stimulation of insulin secretion by pancreatic tissue, suppression of renin production, chemoattraction of leukocytes and initiation of growth-related signaling events, such as activation of oncogenes, protein kinase C, and mitogen-activated protein kinases. 12-lipoxygenase activity and 12-HETE production are also important determining factors in tumor cell metastasis and have been implicated in human prostate cancer and breast cancer (Honn et al., *Cancer Metastasis Rev.*,13:365-396, 1994, Gao et al., *Adv. Exp. Med. Biol.*, 407:41-53, 1997; Natarajan et al., *J. Clin. Endocr. Metab.*, 82:1790-1789, 1997,). Further, 12-HETE has also been implicated in inflammatory skin diseases such as psoriasis (Hussain et al., *Am. J. Physiol.*, 266:243-253, 1994). As mentioned above, metabolism of arachidonic acid by 12-lipoxygenase further generates lipoxins and hepoxillins. Lipoxins play the role of both immunologic and hemodynamic regulators and a variety of biological activities have been reported for hepoxillins which are related to the release of intracellular calcium and the opening of potassium channels (Yamamoto et al., *Pro. Lipid Res.*, 36:23-41, 1997).

The 5-lipoxygenase (5- LO) is perhaps the most important of these enzymes since it leads to the synthesis of leukotrienes. Activation of the 5-LO enzyme involves its docking to a protein termed 5-lipoxygenase-activating protein (FLAP). This binding activates the enzyme, results in its association with the cell membrane and increased synthesis of 5-HPETE and leukotrienes. Leukotriene A (LTA) synthase is associated with 5-lipoxygenase and promotes the rearrangement of 5-HPETE to an unstable

intermediate LTA_4; which may be transformed to LTB_4 by leukotriene A_4 hydrolase (LTA4H); alternatively, it may be conjugated with glutathione by LTC_4 synthase to form LTC_4. LTA4 hydrolase is a pivotal element in leukotriene biosynthesis. Omega-oxidation is regarded as the major pathway for the catabolism of LTB_4. This reaction is catalyzed by LTB_4 omega-hydroxylase (LTB4H3) also called CYP4F2. LTD_4 is produced by the removal of glutamic acid from LTC_4 and LTE_4 results from the subsequent cleavage of glycine; the reincorporation of glutamic acid yields LTF_4.

Epoxygenase pathway

Arachidonate is metabolized to a variety of metabolites by enzymes that contain cytochrome P450. The epoxygenase pathway of the arachidonic acid cascade leads to the formation of epoxyeicosatrienoic acids (EETs) and dihydroxyeicosatrienoic acids (DHETs). CYP2J2 is a human cytochrome P450 arachidonic acid epoxygenase expressed in extrahepatic tissues and particularly in the intestine. In addition to the known effects on intestinal vascular tone, CYP2J2 products may be involved in the release of intestinal neuropeptides, control of intestinal motility and modulation of intestinal fluid/electrolyte transport.

Eicosanoid receptors

The diversity of the effects of eicosanoids is explained by the existence of a number of distinct receptors that mediate their actions. All prostaglandin receptors identified to date are coupled to effector mechanisms through G proteins. Distinct receptors for leukotrienes also have been identified in different tissues, all of these appear to activate phospholipase C.

Therapeutic Agents Interacting with Arachidonic Acid Metabolism

Because of their involvement in so many disease states, there has been a considerable effort to develop effective inhibitors to the formation or action of the eicosanoids. The drugs that influence the eicosanoid pathways are the most commonly used drugs in the world today. Their major uses are to reduce pain, fever and inflammation. Several classes of drugs, most notably the nonsteroidal antiinflammatory drugs (NSAIDs) owe their therapeutic effects to blockade of the formation of eicosanoids. Selective inhibitors of arachidonic acid metabolism also have an important therapeutic value. Inhibition of cyclooxygenase (COX), the enzyme responsible for the biosynthesis of the prostaglandins and certain related autacoids, generally is thought to be a major facet of the mechanism of NSAIDs. Aspirin and newer, widely used drugs belong to the NSAIDs. All NSAIDs are antipyretic, analgesic and antiinflammatory but there are important differences in their activities and in their side effects. The reasons for such differences are not fully understood. Side effects of these drugs include

gastrointestinal ulceration, disturbances in platelet function, changes in renal function and hypersensitivity reactions. It is now appreciated that there are two forms of cyclooxygenase (COX), inhibition of COX-2 is thought to mediate the antipyretic, analgesic and antiinflammatory action of NSAIDs, whereas the simultaneous inhibition of COX-1 may result in unwanted side effects. Efforts are under way to identify COX-2 specific agents. But, it is also possible that enhanced generation of lipoxygenase products, due to the diversion of arachidonic acid metabolism from the cyclooxygenase pathway towards the lipoxygenase pathways, contributes to some of the side effects. Effort is being devoted to a search for drugs that will produce more selective interventions by acting farther along the biosynthetic pathways. Several compounds have been described that selectively antagonize responses to TXA2 and to PGH_2. Some are receptor antagonists others directly inhibit thromboxane synthase.

Advances in understanding the pathobiology of the inflammatory process has suggested several novel approaches for development of drugs to block this process. These include phospholipase A_2 inhibitors. Glucocorticoids are thought to have an effect on arachidonic acid metabolism through the induction of lipocortin that inhibits phospholipase A_2.

NSAIDs generally do not inhibit the formation of other eicosanoids such as the lipoxygenase-produced leukotrienes. Substantial evidence indicates that leukotrienes contribute to the inflammatory response through a variety of effects. Leukotrienes have been implicated as mediators of inflammation and immediate hypersensitivity reactions - in particular, human bronchial asthma - and thus considerable effort has been done to develop either inhibitors of the production or blockers of the action of the actions of these mediators. Various therapeutic approaches have been used including 5-lipoxygenase inhibitors, which block leukotriene formation, or cysteinyl leukotriene receptor antagonists, which block receptor function. LTC_4 synthase is another key step in biosynthesis of leukotrienes and represents another possible site for therapeutic intervention. Drugs targeting leukotriene biosynthesis are being tested and used for their utility in the treatment of various inflammatory conditions.

Most of these drugs are efficacious in providing relief but all available agents have associated, and sometimes severe, toxicity. Certain individuals display intolerance to aspirin and to other drugs acting on arachidonic acid metabolism; this is manifest by symptoms that range from liver toxicity, gastric and intestinal ulceration, disturbance in platelet function, renal injury, nephritis, vasomotor rhinitis with profuse watery secretions, angioneurotic edema, generalized urticaria, and bronchial asthma to laryngeal edema and bronchoconstriction, hypotension, and shock. The underlying mechanism for these severe side effects is not known. Moreover, while these agents have been highly useful for treatment of acute, self-limited inflammatory conditions; their ability to modify disease progression in chronic inflammatory settings remains an area of controversy. The complexity of the

highly regulated pathways and enzymes that lead to the formation of the eicosanoids, has limited the precise identification of the metabolites and enzymes in the arachidonic acid cascade, which play the causal role in pathologies or in side effects to some drugs.

Pharmacogenomics and Arachidonic Acid Metabolism

The vast majority of common diseases, such as cancer, hypertension, diabetes and some inflammatory diseases are polygenic, meaning that they are caused by multiple genes. In addition, these diseases are modulated by environmental factors such as pollutants, chemicals and diet. This is why many diseases are called multifactorial; they result from a synergistic combination of factors, both genetic and environmental. Therapeutic management and drug development could be markedly improved by the identification of specific genetic polymorphisms that determine and predict patient susceptibility to diseases or patient responses to drugs.

To assess the origins of individual variations in disease susceptibility or drug response, pharmacogenomics uses the genomic technologies to identify polymorphisms within genes which are part of biological pathways involved in disease susceptibility, etiology, and development, or more specifically in drug response pathways responsible for a drug's efficacy, tolerance or toxicity. It can provide tools to refine the design of drug development by decreasing the incidence of adverse events in drug tolerance studies, by better defining patient subpopulations of responders and non-responders in efficacy studies and, by combining the results obtained therefrom, to further allow better enlightened individualized drug usage based on efficacy/tolerance prognosis. Pharmacogenomics can also provide tools to identify new targets for designing drugs and to optimize the use of already existing drugs, in order to either increase their response rate and/or exclude non-responders from corresponding treatment, or decrease their undesirable side effects and/or exclude from corresponding treatment patients with marked susceptibility to undesirable side effects. However, for pharmacogenomics to become clinically useful on a large scale, molecular tools and diagnostics tests must become available.

Inflammatory reactions, which are involved in numerous diseases, are highly relevant to pharmacogenomics both because they are at the core of many widespread serious diseases, and because targeting inflammation pathways to design new efficient drugs includes numerous risks of potentiating serious side effects. Arachidonic acid metabolism is particularly relevant since its products, the eicosanoids, are powerful inflammatory molecules and play a role in a number of physiological functions.

Genetic Analysis of Complex Traits

Until recently, the identification of genes linked with detectable traits has relied mainly on a statistical approach called linkage analysis. Linkage analysis is based upon establishing a correlation between the transmission of genetic markers and that of a specific trait throughout generations within a family. Linkage analysis involves the study of families with multiple affected individuals and is useful in the detection of inherited-traits, which are caused by a single gene, or possibly a very small number of genes. Linkage analysis has been successfully applied to map simple genetic traits that show clear Mendelian inheritance patterns and which have a high penetrance (the probability that a person with a given genotype will exhibit a trait). About 100 pathological trait-causing genes have been discovered using linkage analysis over the last 10 years.

But, linkage studies have proven difficult when applied to complex genetic traits. Most traits of medical relevance do not follow simple Mendelian monogenic inheritance. However, complex diseases often aggregate in families, which suggests that there is a genetic component to be found. Such complex traits are often due to the combined action of multiple genes as well as environmental factors. Such complex trait, include susceptibilities to heart disease, hypertension, diabetes, cancer and inflammatory diseases. Drug efficacy, response and tolerance/toxicity can also be considered as multifactoral traits involving a genetic component in the same way as complex diseases. Linkage analysis cannot be applied to the study of such traits for which no large informative families are available. Moreover, because of their low penetrance, such complex traits do not segregate in a clear-cut Mendelian manner as they are passed from one generation to the next. Attempts to map such diseases have been plagued by inconclusive results, demonstrating the need for more sophisticated genetic tools.

Knowledge of genetic variation in the arachidonic acid cascade is important for understanding why some people are more susceptible to disease involving arachidonic acid metabolites or respond differently to treatments targeting arachidonic acid metabolism. Ways to identify genetic polymorphism and to analyze how they impact and predict disease susceptibility and response to treatment are needed.

Although the genes involved in arachidonic acid metabolism represent major drug targets and are of high relevance to pharmaceutical research, we still have scant knowledge concerning the extent and nature of sequence variation in these genes and their regulatory elements. For example, the cDNA and part of the genomic sequence for human 12-lipoxygenase have been cloned and sequenced (Izumi et al., *Proc. Natl. Acad. Sci. USA*, 87:7477-7481, 1990; Funk et al., *Proc. Natl. Acad. Sci. USA*, 87:5638-5642, 1990; Yoshimoto et al., *Biochem. Biophys. Res. Commun.*, 172:1230-1235, 1990, Yoshimoto, et al., *J. Biol. Chem.*, 267:24805-24809, 1992). However, the complete genomic sequence of the 12-lipoxygenase, including its regulatory elements, have not been described.

In the cases where polymorphisms have been identified, the relevance of the variation is rarely understood. While polymorphisms hold promise for use as genetic markers in determining which genes contribute to multigenic or quantitative traits, suitable markers and suitable methods for exploiting those markers have not been found and brought to bare on the genes related to arachidonic acid metabolism.

SUMMARY OF THE INVENTION

The present invention is based on the discovery of a set of novel eicosanoid-related biallelic markers. See Figure 2. These markers are located in the coding regions as well as non-coding regions adjacent to genes which express proteins associated with arachidonic acid metabolism. The position of these markers and knowledge of the surrounding sequence has been used to design polynucleotide compositions which are useful in determining the identity of nucleotides at the marker position, as well as more complex association and haplotyping studies which are useful in determining the genetic basis for disease states involving arachidonic acid metabolism. In addition, the compositions and methods of the invention find use in the identification of the targets for the development of pharmaceutical agents and diagnostic methods, as well as the characterization of the differential efficacious responses to and side effects from pharmaceutical agents acting on arachidonic acid metabolism.

The present invention further stems from the isolation and characterization of the genomic sequence of the 12-lipoxygenase gene including its regulatory regions and of the complete cDNA sequence encoding the 12-lipoxygenase enzyme. Oligonucleotide probes and primers hybridizing specifically with a genomic sequence of 12-lipoxygenase are also part of the invention. Furthermore, an object of the invention consists of recombinant vectors comprising any of the nucleic acid sequences described in the present invention, and in particular of recombinant vectors comprising the promoter region of 12-lipoxygenase or a sequence encoding the 12-lipoxygenase enzyme, as well as cell hosts comprising said nucleic acid sequences or recombinant vectors. The invention also encompasses methods of screening of molecules which, modulate or inhibit the expression of the 12-lipoxygenase gene. The invention is also directed to biallelic markers that are located within the 12-lipoxygenase genomic sequence, these biallelic markers representing useful tools in order to identify a statistically significant association between specific alleles of 12-lipoxygenase gene and one or several disorders related to asthma and/or hepatotoxicity.

A first embodiment of the invention encompasses polynucleotides consisting of, consisting essentially of, or comprising a contiguous span of nucleotides of a sequence selected as an individual or in any combination from the group consisting of SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; the complements thereof; preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-

646, and 651-652, the complements thereof; the sequences described in any one or more of Figure 3, 4, 5, 6, 7, 8, and 9, and the complements thereof, wherein said contiguous span is at least 6, 8, 10, 12, 15, 20, 25, 30, 35, 40, 50, 75, 100, 200, 500, or 1000 nucleotides in length, to the extent that such a length is consistent with the lengths of the particular Sequence ID. The present invention also relates to polynucleotides hybridizing under stringent or intermediate conditions to a sequence selected from the group consisting of SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; the complements thereof; preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652, and the complements thereof. In addition, the polynucleotides of the invention encompass polynucleotides with any further limitation described in this disclosure, or those following, specified alone or in any combination: Said contiguous span may optionally include the eicosanoid-related biallelic marker in said sequence; Optionally either the original or the alternative allele of Figure 4 may be specified as being present at said eicosanoid-related biallelic marker; Optionally either the first or the second allele of Figure 3 or 5 may be specified as being present at said eicosanoid-related biallelic marker; Optionally, said polynucleotide may consists of, or consist essentially of a contiguous span which ranges in length from 8, 10, 12, 15, 18 or 20 to 25, 35, 40, 50, 60, 70, or 80 nucleotides, or be specified as being 12, 15, 18, 20, 25, 35, 40, or 50 nucleotides in length and including an eicosanoid-related biallelic marker of said sequence, and optionally the original allele of Figure 4 is present at said biallelic marker; Optionally, said biallelic marker may be within 6, 5, 4, 3, 2, or 1 nucleotides of the center of said polynucleotide or at the center of said polynucleotide; Optionally, the 3' end of said contiguous span may be present at the 3' end of said polynucleotide; Optionally, biallelic marker may be present at the 3' end of said polynucleotide; Optionally, the 3' end of said polynucleotide may be located within or at least 2, 4, 6, 8, 10, 12, 15, 18, 20, 25, 50, 100, 250, 500, or 1000 nucleotides upstream of an eicosanoid-related biallelic marker in said sequence, to the extent that such a distance is consistent with the lengths of the particular Sequence ID; Optionally, the 3' end of said polynucleotide may be located 1 nucleotide upstream of an eicosanoid-related biallelic marker in said sequence; and Optionally, said polynucleotide may further comprise a label.

A second embodiment of the invention encompasses any polynucleotide of the invention attached to a solid support. In addition, the polynucleotides of the invention which are attached to a solid support encompass polynucleotides with any further limitation described in this disclosure, or those following, specified alone or in any combination: Optionally, said polynucleotides may be specified as attached individually or in groups of at least 2, 5, 8, 10, 12, 15, 20, or 25 distinct polynucleotides of the inventions to a single solid support; Optionally, polynucleotides other than those of the invention may attached to the same solid support as polynucleotides of the invention; Optionally,

when multiple polynucleotides are attached to a solid support they may be attached at random locations, or in an ordered array; Optionally, said ordered array may be addressable.

A third embodiment of the invention encompasses the use of any polynucleotide for, or any polynucleotide for use in, determining the identity of one or more nucleotides at an eicosanoid-related biallelic marker. In addition, the polynucleotides of the invention for use in determining the identity of one or more nucleotides at an eicosanoid-related biallelic marker encompass polynucleotides with any further limitation described in this disclosure, or those following, specified alone or in any combination. Optionally, said eicosanoid-related biallelic marker may be in a sequence selected individually or in any combination from the group consisting of SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; the complements thereof; preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652; and the complements thereof; Optionally, said polynucleotide may comprise a sequence disclosed in the present specification; Optionally, said polynucleotide may consist of, or consist essentially of any polynucleotide described in the present specification; Optionally, said determining may be performed in a hybridization assay, sequencing assay, microsequencing assay, or an enzyme-based mismatch detection assay; Optionally, said polynucleotide may be attached to a solid support, array, or addressable array; Optionally, said polynucleotide may be labeled.

A fourth embodiment of the invention encompasses the use of any polynucleotide for, or any polynucleotide for use in, amplifying a segment of nucleotides comprising an eicosanoid-related biallelic marker. In addition, the polynucleotides of the invention for use in amplifying a segment of nucleotides comprising an eicosanoid-related biallelic marker encompass polynucleotides with any further limitation described in this disclosure, or those following, specified alone or in any combination: Optionally, said eicosanoid-related biallelic marker may be in a sequence selected individually or in any combination from the group consisting of SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; the complements thereof; preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652; and the complements thereof; Optionally, said polynucleotide may comprise a sequence disclosed in the present specification; Optionally, said polynucleotide may consist of, or consist essentially of any polynucleotide described in the present specification; Optionally, said amplifying may be performed by a PCR or LCR. Optionally, said polynucleotide may be attached to a solid support, array, or addressable array. Optionally, said polynucleotide may be labeled.

A fifth embodiment of the invention encompasses methods of genotyping a biological sample comprising determining the identity of a nucleotide at an eicosanoid-related biallelic marker. In addition, the genotyping methods of the invention encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination: Optionally, said eicosanoid-related biallelic marker may be in a sequence selected individually or in any combination

from the group consisting of SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; the complements thereof; preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652; and the complements thereof; Optionally, said method further comprises determining the identity of a second nucleotide at said biallelic marker, wherein said first nucleotide and second nucleotide are not base paired (by Watson & Crick base pairing) to one another; Optionally, said biological sample is derived from a single individual or subject; Optionally, said method is performed *in vitro*; Optionally, said biallelic marker is determined for both copies of said biallelic marker present in said individual's genome; Optionally, said biological sample is derived from multiple subjects or individuals; Optionally, said method further comprises amplifying a portion of said sequence comprising the biallelic marker prior to said determining step; Optionally, wherein said amplifying is performed by PCR, LCR, or replication of a recombinant vector comprising an origin of replication and said portion in a host cell; Optionally, wherein said determining is performed by a hybridization assay, sequencing assay, microsequencing assay, or an enzyme-based mismatch detection assay.

A sixth embodiment of the invention comprises methods of estimating the frequency of an allele in a population comprising genotyping individuals from said population for an eicosanoid-related biallelic marker and determining the proportional representation of said biallelic marker in said population. In addition, the methods of estimating the frequency of an allele in a population of the invention encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination: Optionally, said eicosanoid-related biallelic marker may be in a sequence selected individually or in any combination from the group consisting of SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; the complements thereof; preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652; and the complements thereof; Optionally, determining the frequency of a biallelic marker allele in a population may be accomplished by determining the identity of the nucleotides for both copies of said biallelic marker present in the genome of each individual in said population and calculating the proportional representation of said nucleotide at said eicosanoid-related biallelic marker for the population; Optionally, determining the frequency of a biallelic marker allele in a population may be accomplished by performing a genotyping method on a pooled biological sample derived from a representative number of individuals, or each individual, in said population, and calculating the proportional amount of said nucleotide compared with the total.

A seventh embodiment of the invention comprises methods of detecting an association between an allele and a phenotype, comprising the steps of a) determining the frequency of at least one eicosanoid-related biallelic marker allele in a case population, b) determining the frequency of said eicosanoid-related biallelic marker allele in a control population and; c) determining whether a

statistically significant association exists between said genotype and said phenotype. In addition, the methods of detecting an association between an allele and a phenotype of the invention encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination: Optionally, said eicosanoid-related biallelic marker may be in a sequence selected individually or in any combination from the group consisting of SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; the complements thereof; preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652; and the complements thereof; Optionally, said control population may be a trait negative population, or a random population; Optionally, said phenotype is a disease involving arachidonic acid metabolism, a response to an agent acting on arachidonic acid metabolism, or a side effects to an agent acting on arachidonic acid metabolism; Optionally, the identity of the nucleotides at the biallelic markers in everyone of the following sequences: SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; or preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652 is determined in steps a) and b).

An eighth embodiment of the present invention encompasses methods of estimating the frequency of a haplotype for a set of biallelic markers in a population, comprising the steps of: a) genotyping each individual in said population for at least one eicosanoid-related biallelic marker, b) genotyping each individual in said population for a second biallelic marker by determining the identity of the nucleotides at said second biallelic marker for both copies of said second biallelic marker present in the genome; and c) applying a haplotype determination method to the identities of the nucleotides determined in steps a) and b) to obtain an estimate of said frequency. In addition, the methods of estimating the frequency of a haplotype of the invention encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination: Optionally said haplotype determination method is selected from the group consisting of asymmetric PCR amplification, double PCR amplification of specific alleles, the Clark method, or an expectation maximization algorithm; Optionally, said second biallelic marker is an eicosanoid-related biallelic marker in a sequence selected from the group consisting of the biallelic markers of SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; the complements thereof; preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652; and the complements thereof; Optionally, the identity of the nucleotides at the biallelic markers in everyone of the sequences of SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; or preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652 is determined in steps a) and b).

A ninth embodiment of the present invention encompasses methods of detecting an association between a haplotype and a phenotype, comprising the steps of: a) estimating the frequency of at least one haplotype in a case population according to a method of estimating the frequency of a haplotype of

the invention; b) estimating the frequency of said haplotype in a control population according to the method of estimating the frequency of a haplotype of the invention; and c) determining whether a statistically significant association exists between said haplotype and said phenotype. In addition, the methods of detecting an association between a haplotype and a phenotype of the invention encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination: Optionally, said eicosanoid-related biallelic marker may be in a sequence selected individually or in any combination from the group consisting of SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; the complements thereof; preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652; and the complements thereof; Optionally, said control population may be a trait negative population, or a random population; Optionally, said phenotype is a disease involving arachidonic acid metabolism, a response to an agent acting on arachidonic acid metabolism, or a side effects to an agent acting on arachidonic acid metabolism; Optionally, the identity of the nucleotides at the biallelic markers in everyone of the following sequences: SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; or preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652 is included in the estimating steps a) and b).

A tenth embodiment of the present invention is a method of administering a drug or a treatment comprising the steps of: a) obtaining a nucleic acid sample from an individual; b) determining the identity of the polymorphic base of at least one eicosanoid-related biallelic marker or 12-LO-related biallelic marker which is associated with a positive response to the treatment or the drug; or at least one biallelic eicosanoid-related marker or 12-LO-related biallelic marker which is associated with a negative response to the treatment or the drug; and c) administering the treatment or the drug to the individual if the nucleic acid sample contains said biallelic marker associated with a positive response to the treatment or the drug or if the nucleic acid sample lacks said biallelic marker associated with a negative response to the treatment or the drug. In addition, the methods of the present invention for administering a drug or a treatment encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination: optionally, said eicosanoid-related biallelic marker or 12-LO-related biallelic marker may be in a sequence selected individually or in any combination from the group consisting of SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; the complements thereof; or preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652; and the complements thereof or optionally, the administering step comprises administering the drug or the treatment to the individual if the nucleic acid sample contains said biallelic marker associated with a positive response to the treatment or the drug and the nucleic acid sample lacks said biallelic marker associated with a negative response to the treatment or the drug.

An eleventh embodiment of the present invention is a method of selecting an individual for inclusion in a clinical trial of a treatment or drug comprising the steps of: a) obtaining a nucleic acid sample from an individual; b) determining the identity of the polymorphic base of at least one eicosanoid-related biallelic marker or 12-LO-related biallelic marker which is associated with a positive response to the treatment or the drug, or at least one eicosanoid-related biallelic marker or 12-LO-related biallelic marker which is associated with a negative response to the treatment or the drug in the nucleic acid sample, and c) including the individual in the clinical trial if the nucleic acid sample contains said eicosanoid-related biallelic marker or 12-LO-related biallelic marker associated with a positive response to the treatment or the drug or if the nucleic acid sample lacks said biallelic marker associated with a negative response to the treatment or the drug. In addition, the methods of the present invention for selecting an individual for inclusion in a clinical trial of a treatment or drug encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination: Optionally, said eicosanoid-related biallelic marker or 12-LO-related biallelic marker may be in a sequence selected individually or in any combination from the group consisting of SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; the complements thereof; or preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652; and the complements thereof, optionally, the including step comprises administering the drug or the treatment to the individual if the nucleic acid sample contains said biallelic marker associated with a positive response to the treatment or the drug and the nucleic acid sample lacks said biallelic marker associated with a negative response to the treatment or the drug.

Additional embodiments are set forth in the Detailed Description of the Invention and in the Examples.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A is a diagram of the arachidonic acid metabolism pathway and Figure 1B is a diagram of the lipoxygenase pathway.

Figure 2(A-L) is a chart containing a list of all of the eicosanoid-related biallelic markers for each gene with an indication of the gene for which the marker is in closest physical proximity, an indication of whether the markers have been validated by microsequencing (with a Y indicating that the markers have been validated by microsequencing and an N indicating that it has not), and an indication of the identity and frequency of the least common allele determined by genotyping (with a blank left to indicate that the frequency has not yet been reported for some markers). The frequencies were determined from DNA samples collected from a random US Caucasian population. When the marker

was determined to be homozygous at the particular location for the random US Caucasian population, the homozygous bases were recorded in the "Genotyping Least Common Allele Frequency" column of Figure 2. For example, Seq. ID No. 16 was determined to be homozygous G/G at the biallelic marker position 478 in the US control population, therefore G/G was recorded in the "Genotyping Least Common Allele Frequency" column.

Figures 3(A-H), 4(A-D), and 5 are charts containing lists of the eicosanoid-related biallelic markers. Each marker is described by indicating its SEQ ID, the biallelic marker ID, and the two most common alleles. Figure 3 is a chart containing a list of biallelic markers surrounded by preferred sequences. In the column labeled, "POSITION RANGE OF PREFERRED SEQUENCE" of Figure 3 regions of particularly preferred sequences are listed for each SEQ ID, which contain an eicosanoid-related biallelic marker, as well as particularly preferred regions of sequences that do not contain an eicosanoid-related biallelic marker but, which are in sufficiently close proximity to an eicosanoid-related biallelic marker to be useful as amplification or sequencing primers.

Figure 6(A-B) is a chart listing particular sequences that are useful for designing some of the primers and probes of the invention. Each sequence is described by indicating its Sequence ID and the positions of the first and last nucleotides (position range) of the particular sequence in the Sequence ID.

Figure 7(A-L) is a chart listing microsequencing primers which have been used to genotype eicosanoid-related biallelic markers (indicated by an *) and other preferred microsequencing primers for use in genotyping eicosanoid-related biallelic markers. Each of the primers which falls within the strand of nucleotides included in the Sequence Listing are described by indicating their Sequence ID number and the positions of the first and last nucleotides (position range) of the primers in the Sequence ID. Since the sequences in the Sequence Listing are single stranded and half the possible microsequencing primers are composed of nucleotide sequences from the complementary strand, the primers that are composed of nucleotides in the complementary strand are described by indicating their SEQ ID numbers and the positions of the first and last nucleotides to which they are complementary (complementary position range) in the Sequence ID.

Figure 8(A-L) is a chart listing amplification primers which have been used to amplify polynucleotides containing one or more eicosanoid-related biallelic markers. Each of the primers which falls within the strand of nucleotides included in the Sequence Listing are described by indicating their Sequence ID number and the positions of the first and last nucleotides (position range) of the primers in the Sequence ID. Since the sequences in the Sequence Listing are single stranded and half the possible amplification primers are composed of nucleotide sequences from the complementary strand, the primers that are composed of nucleotides in the complementary strand are defined by the SEQ ID

numbers and the positions of the first and last nucleotides to which they are complementary (complementary position range) in the Sequence ID.

Figure 9(A-F) is a chart listing preferred probes useful in genotyping eicosanoid-related biallelic markers by hybridization assays. The probes are 25-mers with an eicosanoid-related biallelic marker in the center position, and described by indicating their Sequence ID number and the positions of the first and last nucleotides (position range) of the probes in the Sequence ID. The probes complementary to the sequences in each position range in each Sequence ID are also understood to be a part of this preferred list even though they are not specified separately.

Figure 10 is a diagram showing the genomic structure of the FLAP gene and the positions of biallelic markers in close proximity of this gene.

Figure 11 is a graph showing the results of the single point association study between biallelic markers from the FLAP gene and asthma.

Figure 12 is a table showing the results of the association study between biallelic marker haplotypes from the FLAP gene and asthma.

Figure 13 is a table showing the results of the permutation test confirming the statistical significance of the association between asthma and biallelic marker haplotypes from the FLAP gene.

Figure 14 is a diagram showing the genomic structure of the 12-lipoxygenase gene and the positions of biallelic markers in close proximity of this gene.

Figure 15 is a table showing the results of the association study between 12 biallelic marker haplotypes from the 12-LO gene and asthma.

Figure 16A is a table showing the results of allele frequency analysis between 17 12-LO biallelic markers and asthma. Figure 16B is a table showing the results of the association study between 17 12-LO biallelic marker haplotypes from the 12-LO gene and asthma.

Figure 17 is a table showing the results of the association study between 12 biallelic marker haplotypes from the 12-LO gene and hepatotoxicity upon treatment with zileuton.

Figure 18A is a table showing the results of the allele frequency analysis between 17 12-LO biallelic markers and hepatotoxicity upon treatment with zileuton. Figure 18B is a table showing the results of the association study between 17 12-LO biallelic marker haplotypes from the 12-LO gene and hepatotoxicity upon treatment with zileuton.

Figure 19 is a table showing a summary of the association study results, permutation tests confirming the statistical significance of the association between asthma and biallelic marker haplotypes from the 12-LO gene, and permutation tests confirming the statistical significance of the association between secondary effects upon treatment with zileuton and biallelic marker haplotypes from the 12-LO gene.

Figure 20 is a table showing a summary of the association study results, permutation tests confirming the statistical significance of the association between asthma and additional biallelic marker haplotypes from the 12-LO gene, and permutation tests confirming the statistical significance of the association between secondary effects upon treatment with zileuton and biallelic marker haplotypes from the 12-LO gene.

Figure 21 is a chart containing a list of preferred 12-LO-related biallelic markers with an indication of the frequency of the least common allele determined by genotyping. Frequencies were determined in a random US Caucasian population, in an asthmatic population showing no side effects upon treatment with Zyflo™ (ALT-) and in an asthmatic population showing elevated alanine aminotransferase levels upon treatment with Zyflo™ (ALT+).

Figure 22 is a block diagram of an exemplary computer system.

Figure 23 is a flow diagram illustrating one embodiment of a process 200 for comparing a new nucleotide or protein sequence with a database of sequences in order to determine the homology levels between the new sequence and the sequences in the database.

Figure 24 is a flow diagram illustrating one embodiment of a process 250 in a computer for determining whether two sequences are homologous.

Figure 25 is a flow diagram illustrating one embodiment of an identifier process 300 for detecting the presence of a feature in a sequence.

DETAILED DESCRIPTION OF THE INVENTION

Advantages of the biallelic markers of the present invention

The eicosanoid-related biallelic markers of the present invention offer a number of important advantages over other genetic markers such as RFLP (Restriction fragment length polymorphism) and VNTR (Variable Number of Tandem Repeats) markers.

The first generation of markers, were RFLPs, which are variations that modify the length of a restriction fragment. But methods used to identify and to type RFLPs are relatively wasteful of materials, effort, and time. The second generation of genetic markers were VNTRs, which can be categorized as either minisatellites or microsatellites. Minisatellites are tandemly repeated DNA sequences present in units of 5-50 repeats which are distributed along regions of the human chromosomes ranging from 0.1 to 20 kilobases in length. Since they present many possible alleles, their informative content is very high. Minisatellites are scored by performing Southern blots to identify the number of tandem repeats present in a nucleic acid sample from the individual being tested. However, there are only 10^4 potential VNTRs

that can be typed by Southern blotting. Moreover, both RFLP and VNTR markers are costly and time-consuming to develop and assay in large numbers.

Single nucleotide polymorphism or biallelic markers can be used in the same manner as RFLPs and VNTRs but offer several advantages. Single nucleotide polymorphisms are densely spaced in the human genome and represent the most frequent type of variation. An estimated number of more than 10^7 sites are scattered along the $3x10^9$ base pairs of the human genome. Therefore, single nucleotide polymorphism occur at a greater frequency and with greater uniformity than RFLP or VNTR markers which means that there is a greater probability that such a marker will be found in close proximity to a genetic locus of interest. Single nucleotide polymorphisms are less variable than VNTR markers but are mutationally more stable.

Also, the different forms of a characterized single nucleotide polymorphism, such as the biallelic markers of the present invention, are often easier to distinguish and can therefore be typed easily on a routine basis. Biallelic markers have single nucleotide based alleles and they have only two common alleles, which allows highly parallel detection and automated scoring. The biallelic markers of the present invention offer the possibility of rapid, high-throughput genotyping of a large number of individuals.

Biallelic markers are densely spaced in the genome, sufficiently informative and can be assayed in large numbers. The combined effects of these advantages make biallelic markers extremely valuable in genetic studies. Biallelic markers can be used in linkage studies in families, in allele sharing methods, in linkage disequilibrium studies in populations, in association studies of case-control populations. An important aspect of the present invention is that biallelic markers allow association studies to be performed to identify genes involved in complex traits. Association studies examine the frequency of marker alleles in unrelated case- and control-populations and are generally employed in the detection of polygenic or sporadic traits. Association studies may be conducted within the general population and are not limited to studies performed on related individuals in affected families (linkage studies). Biallelic markers in different genes can be screened in parallel for direct association with disease or response to a treatment. This multiple gene approach is a powerful tool for a variety of human genetic studies as it provides the necessary statistical power to examine the synergistic effect of multiple genetic factors on a particular phenotype, drug response, sporadic trait, or disease state with a complex genetic etiology.

Candidate genes of the present invention

Different approaches can be employed to perform association studies: genome-wide association studies, candidate region association studies and candidate gene association studies. Genome-wide

association studies rely on the screening of genetic markers evenly spaced and covering the entire genome. Candidate region association studies rely on the screening of genetic markers evenly spaced covering a region identified as linked to the trait of interest. The candidate gene approach is based on the study of genetic markers specifically derived from genes potentially involved in a biological pathway related to the trait of interest. In the present invention, genes involved in arachidonic acid metabolism have been chosen as candidate genes. This metabolic pathway leads to the biosynthesis of eicosanoids, which are chemical mediators that play an important role in a number of inflammatory diseases, moreover, these pathways are important drug targets and genetic polymorphisms in these genes are highly relevant in the response to a number of drugs. The candidate gene analysis clearly provides a short-cut approach to the identification of genes and gene polymorphisms related to a particular trait when some information concerning the biology of the trait is available as is the case for arachidonic acid metabolism. However, it should be noted that all of the biallelic markers disclosed in the instant application can be employed as part of genome-wide association studies or as part of candidate region association studies and such uses are specifically contemplated in the present invention and claims. All of the markers are known to be in close proximity to the genes with which they are listed in Figure 2. For a portion of the markers, the precise position of the marker with respect to the various coding and non-coding elements of the genes has also been determined.

Definitions

As used interchangeably herein, the terms "oligonucleotides", "nucleic acids" and "polynucleotides" include RNA, DNA, or RNA/DNA hybrid sequences of more than one nucleotide in either single chain or duplex form. The term "nucleotide" as used herein as an adjective to describe molecules comprising RNA, DNA, or RNA/DNA hybrid sequences of any length in single-stranded or duplex form. The term "nucleotide" is also used herein as a noun to refer to individual nucleotides or varieties of nucleotides, meaning a molecule, or individual unit in a larger nucleic acid molecule, comprising a purine or pyrimidine, a ribose or deoxyribose sugar moiety, and a phosphate group, or phosphodiester linkage in the case of nucleotides within an oligonucleotide or polynucleotide. Although the term "nucleotide" is also used herein to encompass "modified nucleotides" which comprise at least one modifications (a) an alternative linking group, (b) an analogous form of purine, (c) an analogous form of pyrimidine, or (d) an analogous sugar, for examples of analogous linking groups, purine, pyrimidines, and sugars see for example PCT publication No. WO 95/04064. However, the polynucleotides of the invention are preferably comprised of greater than 50% conventional deoxyribose nucleotides, and most preferably greater than 90% conventional deoxyribose nucleotides. The polynucleotide sequences of the invention may be prepared by any known method, including

synthetic, recombinant, *ex vivo* generation, or a combination thereof, as well as utilizing any purification methods known in the art.

Throughout the present specification, the expression "nucleotide sequence" may be employed to designate indifferently a polynucleotide or a nucleic acid. More precisely, the expression "nucleotide sequence" encompasses the nucleic material itself and is thus not restricted to the sequence information (i.e. the succession of letters chosen among the four base letters) that biochemically characterizes a specific DNA or RNA molecule.

The term "polypeptide" refers to a polymer of amino without regard to the length of the polymer; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. This term also does not specify or exclude prost-expression modifications of polypeptides, for example, polypeptides which include the covalent attachment of glycosyl groups, acetyl groups, phosphate groups, lipid groups and the like are expressly encompassed by the term polypeptide. Also included within the definition are polypeptides which contain one or more analogs of an amino acid (including, for example, non-naturally occurring amino acids, amino acids which only occur naturally in an unrelated biological system, modified amino acids from mammalian systems etc.), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

The term "recombinant polypeptide" is used herein to refer to polypeptides that have been artificially designed and which comprise at least two polypeptide sequences that are not found as contiguous polypeptide sequences in their initial natural environment, or to refer to polypeptides which have been expressed from a recombinant polynucleotide.

The term "isolated" requires that the material be removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or DNA or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotide could be part of a vector and/or such polynucleotide or polypeptide could be part of a composition, and still be isolated in that the vector or composition is not part of its natural environment.

The term "purified" does not require absolute purity; rather, it is intended as a relative definition. Purification of starting material or natural material to at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated. As an example, purification from 0.1 % concentration to 10 % concentration is two orders of magnitude. The term purified is used herein to describe a polynucleotide or polynucleotide vector of the invention which has been separated from other compounds including, but not limited to

other nucleic acids, carbohydrates, lipids and proteins (such as the enzymes used in the synthesis of the polynucleotide), or the separation of covalently closed polynucleotides from linear polynucleotides. A polynucleotide is substantially pure when at least about 50%, preferably 60 to 75% of a sample exhibits a single polynucleotide sequence and conformation (linear versus covalently close). A substantially pure polynucleotide typically comprises about 50%, preferably 60 to 90% weight/weight of a nucleic acid sample, more usually about 95%, and preferably is over about 99% pure. Polynucleotide purity or homogeneity is indicated by a number of means well known in the art, such as agarose or polyacrylamide gel electrophoresis of a sample, followed by visualizing a single polynucleotide band upon staining the gel. For certain purposes, higher resolution can be provided by using HPLC or other means well known in the art.

The term "primer" denotes a specific oligonucleotide sequence which is complementary to a target nucleotide sequence and used to hybridize to the target nucleotide sequence. A primer serves as an initiation point for nucleotide polymerization catalyzed by DNA polymerase, RNA polymerase or reverse transcriptase.

The term "probe" denotes a defined nucleic acid segment (or nucleotide analog segment, e.g., polynucleotide as defined herein) which can be used to identify a specific polynucleotide sequence present in samples, said nucleic acid segment comprising a nucleotide sequence complementary of the specific polynucleotide sequence to be identified.

The term "disease involving arachidonic acid metabolism" refers to a condition linked to disturbances in expression, production or cellular response to eicosanoids such as prostaglandins, thromboxanes, prostacyclins, leukotrienes or hydroperoxyeicosaetrenoic acids. A disease involving arachidonic acid metabolism further refers to a condition involving one or several enzymes of the distinct enzyme systems contributing to arachidonate metabolism including particularly the cyclooxygenase pathway and the lipoxygenase pathway and the arachadonic acid metabolites of such systems including 12-HETE, 12-HPETE, lipoxins and hepoxolins. "Diseases involving arachidonic acid metabolism" also include chronic inflammatory diseases, acute allergic inflammation and inflammatory conditions such as pain, fever, hypersensitivity, asthma, psoriasis and arthritis. "Diseases involving arachidonic acid metabolism" also include disorders in platelet function, blood pressure, thrombosis, renal function, host defense mechanism, hemostasis, smooth muscle tone, male infertility, primary dysmenorrhea, disorders in parturition, and disorders in tissue injury repair, as well as disorders in cellular function and development. "Diseases involving arachidonic acid metabolism" also include diseases such as gastrointestinal ulceration, coronary and cerebrovascular syndromes, glomerular immune injury and cancer.

The term "agent acting on arachidonic acid metabolism" refers to a drug or a compound modulating the activity or concentration of an enzyme or regulatory molecule involved in arachidonic acid metabolism, including but not limited to cyclooxygenase, prostacyclin synthase, thromboxane synthase, lipoxygenases, 5-lipoxygenase and 5-lipoxygenase activating protein. "Agent acting on arachidonic acid metabolism" further refers to non-steroidal antiinflammatory drugs (NSAIDs), eicosanoid receptor antagonists, eicosanoid analogs, COX-1 inhibitors, COX-2 inhibitors, thromboxane synthase inhibitors, 5-lipoxygenase inhibitors and 5-lipoxygenase activating protein inhibitors. "Agent acting on arachidonic acid metabolism" also refers to compounds modulating the formation and action of eicosanoids such as prostaglandins, prostacyclins, thromboxanes, leukotrienes or hydroperoxyeicosaetrenoic acids.

The terms "response to an agent acting on arachidonic acid metabolism" refer to drug efficacy, including but not limited to ability to metabolize a compound, to the ability to convert a pro-drug to an active drug, and to the pharmacokinetics (absorption, distribution, elimination) and the pharmacodynamics (receptor-related) of a drug in an individual.

The terms "side effects to an agent acting on arachidonic acid metabolism" refer to adverse effects of therapy resulting from extensions of the principal pharmacological action of the drug or to idiosyncratic adverse reactions resulting from an interaction of the drug with unique host factors. "Side effects to an agent acting on arachidonic acid metabolism" include, but are not limited to, adverse reactions such as dermatologic, hematologic or hepatologic toxicities and further includes gastric and intestinal ulceration, disturbance in platelet function, renal injury, nephritis, vasomotor rhinitis with profuse watery secretions, angioneurotic edema, generalized urticaria, and bronchial asthma to laryngeal edema and bronchoconstriction, hypotension, and shock.

The terms "trait" and "phenotype" are used interchangeably herein and refer to any visible, detectable or otherwise measurable property of an organism such as symptoms of, or susceptibility to a disease for example. Typically the terms "trait" or "phenotype" are used herein to refer to symptoms of, or susceptibility to a disease involving arachidonic acid metabolism; or to refer to an individual's response to an agent acting on arachidonic acid metabolism; or to refer to symptoms of, or susceptibility to side effects to an agent acting on arachidonic acid metabolism.

The terms "agent acting on 5-lipoxygenase" refers to a drug or a compound modulating the activity or concentration of the 5-lipoxygenase enzyme such as 5-lipoxygenase inhibitors. "Agent acting on 5-lipoxygenase" also refers to compounds modulating the formation and action of leukotrienes.

The terms "side effects to an agent acting on 5-lipoxygenase" include, but are not limited to, adverse reactions such as dermatologic, hematologic or hepatologic toxicities.

The term "allele" is used herein to refer to variants of a nucleotide sequence. A biallelic polymorphism has two forms. Typically the first identified allele is designated as the original allele whereas other alleles are designated as alternative alleles. Diploid organisms may be homozygous or heterozygous for an allelic form.

The term "heterozygosity rate" is used herein to refer to the incidence of individuals in a population, which are heterozygous at a particular allele. In a biallelic system the heterozygosity rate is on average equal to $2P_a(1-P_a)$, where P_a is the frequency of the least common allele. In order to be useful in genetic studies a genetic marker should have an adequate level of heterozygosity to allow a reasonable probability that a randomly selected person will be heterozygous.

The term "genotype" as used herein refers the identity of the alleles present in an individual or a sample. In the context of the present invention a genotype preferably refers to the description of the biallelic marker alleles present in an individual or a sample. The term "genotyping" a sample or an individual for a biallelic marker consists of determining the specific allele or the specific nucleotide carried by an individual at a biallelic marker.

The term "mutation" as used herein refers to a difference in DNA sequence between or among different genomes or individuals which has a frequency below 1%.

The term "haplotype" refers to a combination of alleles present in an individual or a sample. In the context of the present invention a haplotype preferably refers to a combination of biallelic marker alleles found in a given individual and which may be associated with a phenotype.

The term "polymorphism" as used herein refers to the occurrence of two or more alternative genomic sequences or alleles between or among different genomes or individuals. "Polymorphic" refers to the condition in which two or more variants of a specific genomic sequence can be found in a population. A "polymorphic site" is the locus at which the variation occurs. A single nucleotide polymorphism is a single base pair change. Typically a single nucleotide polymorphism is the replacement of one nucleotide by another nucleotide at the polymorphic site. Deletion of a single nucleotide or insertion of a single nucleotide, also give rise to single nucleotide polymorphisms. In the context of the present invention "single nucleotide polymorphism" preferably refers to a single nucleotide substitution. Typically, between different genomes or between different individuals, the polymorphic site may be occupied by two different nucleotides.

The terms "biallelic polymorphism" and "biallelic marker" are used interchangeably herein to refer to a polymorphism having two alleles at a fairly high frequency in the population, preferably a single nucleotide polymorphism. A "biallelic marker allele" refers to the nucleotide variants present at a biallelic marker site. Typically the frequency of the less common allele of the biallelic markers of the present invention has been validated to be greater than 1%, preferably the frequency is greater than

10%, more preferably the frequency is at least 20% (i.e. heterozygosity rate of at least 0.32), even more preferably the frequency is at least 30% (i.e. heterozygosity rate of at least 0.42). A biallelic marker wherein the frequency of the less common allele is 30% or more is termed a "high quality biallelic marker."

The location of nucleotides in a polynucleotide with respect to the center of the polynucleotide are described herein in the following manner. When a polynucleotide has an odd number of nucleotides, the nucleotide at an equal distance from the 3' and 5' ends of the polynucleotide is considered to be "at the center" of the polynucleotide, and any nucleotide immediately adjacent to the nucleotide at the center, or the nucleotide at the center itself is considered to be "within 1 nucleotide of the center." With an odd number of nucleotides in a polynucleotide any of the five nucleotides positions in the middle of the polynucleotide would be considered to be within 2 nucleotides of the center, and so on. When a polynucleotide has an even number of nucleotides, there would be a bond and not a nucleotide at the center of the polynucleotide. Thus, either of the two central nucleotides would be considered to be "within 1 nucleotide of the center" and any of the four nucleotides in the middle of the polynucleotide would be considered to be "within 2 nucleotides of the center", and so on. For polymorphisms which involve the substitution, insertion or deletion of 1 or more nucleotides, the polymorphism, allele or biallelic marker is "at the center" of a polynucleotide if the difference between the distance from the substituted, inserted, or deleted polynucleotides of the polymorphism and the 3' end of the polynucleotide, and the distance from the substituted, inserted, or deleted polynucleotides of the polymorphism and the 5' end of the polynucleotide is zero or one nucleotide. If this difference is 0 to 3, then the polymorphism is considered to be "within 1 nucleotide of the center." If the difference is 0 to 5, the polymorphism is considered to be "within 2 nucleotides of the center." If the difference is 0 to 7, the polymorphism is considered to be "within 3 nucleotides of the center," and so on. For polymorphisms which involve the substitution, insertion or deletion of 1 or more nucleotides, the polymorphism, allele or biallelic marker is "at the center" of a polynucleotide if the difference between the distance from the substituted, inserted, or deleted polynucleotides of the polymorphism and the 3' end of the polynucleotide, and the distance from the substituted, inserted, or deleted polynucleotides of the polymorphism and the 5' end of the polynucleotide is zero or one nucleotide. If this difference is 0 to 3, then the polymorphism is considered to be "within 1 nucleotide of the center." If the difference is 0 to 5, the polymorphism is considered to be "within 2 nucleotides of the center." If the difference is 0 to 7, the polymorphism is considered to be "within 3 nucleotides of the center," and so on.

A "promoter" refers to a DNA sequence recognized by the synthetic machinery of the cell required to initiate the specific transcription of a gene.

As used herein, the term "operably linked" refers to a linkage of polynucleotide elements in a functional relationship. For instance, a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the coding sequence. More precisely, two DNA molecules (such as a polynucleotide containing a promoter region and a polynucleotide encoding a desired polypeptide or polynucleotide) are said to be "operably linked" if the nature of the linkage between the two polynucleotides does not (1) result in the introduction of a frame-shift mutation or (2) interfere with the ability of the polynucleotide containing the promoter to direct the transcription of the coding polynucleotide.

The term "upstream" is used herein to refer to a location, which is toward the 5' end of the polynucleotide from a specific reference point.

The terms "base paired" and "Watson & Crick base paired" are used interchangeably herein to refer to nucleotides which can be hydrogen bonded to one another be virtue of their sequence identities in a manner like that found in double-helical DNA with thymine or uracil residues linked to adenine residues by two hydrogen bonds and cytosine and guanine residues linked by three hydrogen bonds (See Stryer, L., *Biochemistry,* 4th edition, 1995).

The terms "complementary" or "complement thereof" are used herein to refer to the sequences of polynucleotides which is capable of forming Watson & Crick base pairing with another specified polynucleotide throughout the entirety of the complementary region. This term is applied to pairs of polynucleotides based solely upon their sequences and not any particular set of conditions under which the two polynucleotides would actually bind.

As used herein, the term "non-human animal" refers to any non-human vertebrate, birds and more usually mammals, preferably primates, farm animals such as swine, goats, sheep, donkeys, and horses, rabbits or rodents, more preferably rats or mice. As used herein, the term "animal" is used to refer to any vertebrate, preferable a mammal. Both the terms "animal" and "mammal" expressly embrace human subjects unless preceded with the term "non-human."

As used herein, the term "antibody" refers to a polypeptide or group of polypeptides which are comprised of at least one binding domain, where an antibody binding domain is formed from the folding of variable domains of an antibody molecule to form three-dimensional binding spaces with an internal surface shape and charge distribution complementary to the features of an antigenic determinant of an antigen., which allows an immunological reaction with the antigen. Antibodies include recombinant proteins comprising the binding domains, as wells as fragments, including Fab, Fab', $F(ab)_2$, and $F(ab')_2$ fragments.

As used herein, an "antigenic determinant" is the portion of an antigen molecule, in this case a 12-LO polypeptide, that determines the specificity of the antigen-antibody reaction. An "epitope"

refers to an antigenic determinant of a polypeptide. An epitope can comprise as few as 3 amino acids in a spatial conformation which, is unique to the epitope. Generally an epitope consists of at least 6 such amino acids, and more usually at least 8-10 such amino acids. Methods for determining the amino acids which make up an epitope include x-ray crystallography, 2-dimensional nuclear magnetic resonance, and epitope mapping e.g. the Pepscan method described by H. Mario Geysen et al. 1984. Proc. Natl. Acad. Sci. U.S.A. 81:3998-4002; PCT Publication No. WO 84/03564; and PCT Publication No. WO 84/03506.

As used herein the term "eicosanoid-related biallelic marker" relates to a set of biallelic markers in linkage disequilibrium with all of the genes disclosed in Figure 2 with the exception of FLAP. All of these genes express proteins that are related to eicosanoid metabolism. The term eicosanoid-related biallelic marker encompasses all of the biallelic markers disclosed in Figure 2, preferably the biallelic markers found in SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; or more preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652. The preferred eicosanoid-related biallelic marker alleles of the present invention include each one the alleles described in Figures 2, 3, 4, and 5 individually or in groups consisting of all the possible combinations of the alleles included in Figures 2, 3, 4, and 5, preferably the biallelic markers found in SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; or more preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652.

As used herein the term "12-LO-related biallelic marker" and "12-lipoxygenase-related biallelic marker" are used interchangeably herein to relate to all biallelic markers in linkage disequilibrium with the biallelic markers of the 12- lipoxygenase gene. The term 12-LO-related biallelic marker includes both the genic and non-genic biallelic markers described in Table I.

The term "non-genic" is used herein to describe 12-LO-related biallelic markers, as well as polynucleotides and primers which occur outside the nucleotide positions shown in the human 12-LO genomic sequence of SEQ ID No. 651. The term "genic" is used herein to describe 12-LO-related biallelic markers as well as polynucleotides and primers which do occur in the nucleotide positions shown in the human 12-LO genomic sequence of SEQ ID No1.

The term "sequence described in Figure 3" is used herein to refer to the entire collection of nucleotide sequences or any individual sequence defined in Figure 3. The SEQ ID that contains each "sequence described in Figure 3" is provided in the column labeled, "SEQ ID NO." The range of nucleotide positions within the Sequence ID of which each sequence consists is provided in the same row as the Sequence ID in a column labeled, "POSITION RANGE OF PREFERRED SEQUENCE". It should be noted that some of the Sequence ID numbers have multiple sequence ranges listed, because they contain multiple "sequences described in Figure 3." Unless otherwise noted the term "sequence

described in Figure 3" is to be construed as encompassing sequences that contain either of the two alleles listed in the columns labeled, "1ST ALLELE" and "2ND ALLELE" at the position identified in field <222> of the allele feature in the appended Sequence Listing for each Sequence ID number referenced in Figure 3. For all inventions which relate to biallelic markers or sequences described in Figure 3, a preferred set of markers or sequences excludes Sequence ID Nos. 1-10, 19, 23-25, and 647-650.

The term "sequence described in Figure 4" is used herein to refer to the entire collection of nucleotide sequences or any individual sequence defined in Figure 4. Unless otherwise noted, the "sequences described in Figure 4" consist of the entire sequence of each Sequence ID provided in the column labeled, "SEQ ID NO." Also unless otherwise noted the term "sequence described in Figure 4" is to be construed as encompassing sequences that contain either of the two alleles listed in the columns labeled, "ORIGINAL ALLELE" and "ALTERNATIVE ALLELE" at the position identified in field <222> of the allele feature in the appended Sequence Listing for each Sequence ID number referenced in Figure 4. For all inventions which relate to biallelic markers or sequences described in Figure 4, a preferred set of markers or sequences excludes Sequence ID Nos. 11-18 and 20-21.

The term "sequence described in Figure 5" is used herein to refer to the entire collection of nucleotide sequences or any individual sequence defined in Figure 5. Unless otherwise noted, the "sequences described in Figure 5" consist of the entire sequence of each Sequence ID provided in the column labeled, "SEQ ID NO." Also unless otherwise noted the term "sequence described in Figure 5" is to be construed as encompassing sequences that contain either of the two alleles listed in the columns labeled, "1ST ALLELE" and "2ND ALLELE" at the position identified in field <222> of the allele feature in the appended Sequence Listing for each Sequence ID number referenced in Figure 5. For all inventions which relate to biallelic markers or sequences described in Figure 3, a preferred set of markers or sequences excludes Sequence ID No. 22.

The term "sequence described in Figure 6" is used herein to refer to the entire collection of nucleotide sequences or any individual sequence defined in Figure 6. The SEQ ID that contains each "sequence described in Figure 6" is provided in the column labeled, "SEQ ID NO." The range of nucleotide positions within the Sequence ID of which each sequence consists is provided in the same row as the Sequence ID in a column labeled, "POSITION RANGE OF PREFERRED SEQUENCE". It should be noted that some of the Sequence ID numbers have multiple sequence ranges listed, because they contain multiple "sequences described in Figure 6."

The term "sequence described in Figure 7" is used herein to refer to the entire collection of nucleotide sequences or any individual sequence defined in Figure 7. The SEQ ID that contains each "sequence described in Figure 7" is provided in the column labeled "SEQ ID." The range of nucleotide

positions within the Sequence ID of which half of the sequences consists is provided in the same row as the Sequence ID in a column labeled, "POSITION RANGE OF MICROSEQUENCING PRIMERS." The remaining half of the sequences described in Figure 7 are complementary to the range of nucleotide positions within the Sequence ID provided in the same row as the Sequence ID in a column labeled, "COMPLEMENTARY POSITION RANGE OF MICROSEQUENCING PRIMERS." For all inventions which relate to biallelic markers or sequences described in Figure 7, a more preferred set of markers or sequences consists of those markers or sequences found in SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652.

The term "sequence described in Figure 8" is used herein to refer to the entire collection of nucleotide sequences or any individual sequence defined in Figure 8. The SEQ ID that contains each "sequence described in Figure 8" is provided in the column labeled, "SEQ ID." The range of nucleotide positions within the Sequence ID of which half of the sequences consists is provided in the same row as the Sequence ID in a column labeled, "POSITION RANGE OF AMPLIFICATION PRIMERS." The remaining half of the sequences described in Figure 8 are complementary to the range of nucleotide positions within the Sequence ID provided in the same row as the Sequence ID in a column labeled, "COMPLEMENTARY POSITION RANGE OF AMPLIFICATION PRIMERS." For all inventions which relate to biallelic markers or sequences described in Figure 8, a more preferred set of markers or sequences consists of those markers or sequences found in SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652.

The term "sequence described in Figure 9" is used herein to refer to the entire collection of nucleotide sequences or any individual sequence defined in Figure 9. The SEQ ID that contains each "sequence described in Figure 9" is provided in the column labeled, "SEQ ID". The range of nucleotide positions within the Sequence ID of which each sequence consists is provided in the same row as the Sequence ID in a column labeled, "POSITION RANGE OF PROBES". The sequences which are complementary to the ranges listed in the column labeled, "POSITION RANGE OF PROBES" are also encompassed by the term, "sequence described in Figure 9." Unless otherwise noted the term "sequence described in Figure 9" is to be construed as encompassing sequences that contain either of the two alleles listed in the allele feature in the appended Sequence Listing for each Sequence ID number referenced in Figure 9. For all inventions which relate to biallelic markers or sequences described in Figure 9, a more preferred set of markers or sequences consists of those markers or sequences found in SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652.

The terms "biallelic marker described in Figure" and "allele described in Figure" are used herein to refer to any or all alleles which are listed in the allele feature in the appended Sequence Listing for each Sequence ID number referenced in the particular Figure being mentioned.

The following abbreviations are used in this disclosure: the LTB_4H_2 gene is abbreviated LTB4H2; leukotriene B_4-12-OH dehydrogenase is abbreviated LTB4-12OH; leukotriene B_4 receptor is abbreviated LTB4R; PGD-synthase is abbreviated PGDS; and PG-15-OH dehydrogenase is abbreviated PG15OH.

Variants and Fragments

The invention also relates to variants and fragments of the polynucleotides described herein, particularly of a 12-LO gene containing one or more biallelic markers according to the invention.

Variants of polynucleotides, as the term is used herein, are polynucleotides that differ from a reference polynucleotide. A variant of a polynucleotide may be a naturally occurring variant such as a naturally occurring allelic variant, or it may be a variant that is not known to occur naturally. Such non-naturally occurring variants of the polynucleotide may be made by mutagenesis techniques, including those applied to polynucleotides, cells or organisms. Generally, differences are limited so that the nucleotide sequences of the reference and the variant are closely similar overall and, in many regions, identical. Variants of polynucleotides according to the invention include, without being limited to, nucleotide sequences which are at least 95% identical , preferably at least 99% identical, more particularly at least 99.5% identical, and most preferably at least 99.8% identical to a polynucleotide selected from the group consisting of the polynucleotides of a sequence from any sequence in the Sequence Listing as well as sequences which are complementary thereto or to any polynucleotide fragment of at least 8 consecutive nucleotides of a sequence from any sequence in the Sequence Listing. Nucleotide changes present in a variant polynucleotide may be silent, which means that they do not alter the amino acids encoded by the polynucleotide. However, nucleotide changes may also result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence. The substitutions, deletions or additions may involve one or more nucleotides. The variants may be altered in coding or non-coding regions or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or additions. In the context of the present invention, particularly preferred embodiments are those in which the polynucleotides encode polypeptides which retain substantially the same biological function or activity as the mature 12-LO protein, or those in which the polynucleotides encode polypeptides which maintain or increase a particular biological activity, while reducing a second biological activity. A polynucleotide fragment is a polynucleotide having a sequence that is entirely the same as part but not all of a given nucleotide sequence, preferably the nucleotide sequence of a 12-LO gene, and variants thereof. The fragment can be a portion of an exon or of an intron of a 12-LO gene. It can also be a portion of the regulatory regions of the 12-LO gene preferably of the promoter sequence of the 12-LO

gene. Such fragments may be "free-standing", i.e. not part of or fused to other polynucleotides, or they may be comprised within a single larger polynucleotide of which they form a part or region. Indeed, several of these fragments may be present within a single larger polynucleotide.

Identity between Nucleic Acids and Polypeptides

The terms "percentage of sequence identity" and "percentage homology" are used interchangeably herein to refer to comparisons among polynucleotides and polypeptides, and are determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Homology is evaluated using any of the variety of sequence comparison algorithms and programs known in the art. Such algorithms and programs include, but are by no means limited to, TBLASTN, BLASTP, FASTA, TFASTA, and CLUSTALW (Pearson and Lipman, *Proc. Natl. Acad. Sci.* 85(8):2444-2448, 1988; Altschul et al., *J. Mol. Biol.* 215(3):403-410,1990; Thompson et al., *Nucleic Acids Res.* 22(2):4673-4680, 1994; Higgins et al., *Methods Enzymol.* 266:383-402, 1996; Altschul et al., *Nature Genetics* 3:266-272, 1993). In a particularly preferred embodiment, protein and nucleic acid sequence homologies are evaluated using the Basic Local Alignment Search Tool ("BLAST") which is well known in the art (See, e.g., Karlin and Altschul,. *Proc. Natl. Acad. Sci. USA* 87:2267-2268, 1990; Altschul et al., *J. Mol. Biol.* 215(3):403-410, 1990; Altschul et al., *Nature Genetics* 3:266-272, 1993; Altschul et al., *Nuc. Acids Res.* 25:3389-3402, 1997). In particular, five specific BLAST programs are used to perform the following task:

(1) BLASTP and BLAST3 compare an amino acid query sequence against a protein sequence database;

(2) BLASTN compares a nucleotide query sequence against a nucleotide sequence database;

(3) BLASTX compares the six-frame conceptual translation products of a query nucleotide sequence (both strands) against a protein sequence database;

(4) TBLASTN compares a query protein sequence against a nucleotide sequence database translated in all six reading frames (both strands); and

(5) TBLASTX compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

The BLAST programs identify homologous sequences by identifying similar segments, which are referred to herein as "high-scoring segment pairs," between a query amino or nucleic acid sequence and a test sequence which is preferably obtained from a protein or nucleic acid sequence database. High-scoring segment pairs are preferably identified (i.e., aligned) by means of a scoring matrix, many of which are known in the art. Preferably, the scoring matrix used is the BLOSUM62 matrix (Gonnet et al., *Science* 256:1443-1445, 1992; Henikoff and Henikoff, *Proteins* 17:49-61, 1993). Less preferably, the PAM or PAM250 matrices may also be used (See, e.g., Schwartz and Dayhoff, eds., *Matrices for Detecting Distance Relationships: Atlas of Protein Sequence and Structure, Washington:National Biomedical Research Foundation*, 1978). The BLAST programs evaluate the statistical significance of all high-scoring segment pairs identified, and preferably selects those segments which satisfy a user-specified threshold of significance, such as a user-specified percent homology. Preferably, the statistical significance of a high-scoring segment pair is evaluated using the statistical significance formula of Karlin (see, e.g., Karlin and Altschul, *Proc. Natl. Acad. Sci. USA* 87:2267-2268, 1990).

Stringent Hybridization Conditions

By way of example and not limitation, procedures using conditions of high stringency are as follows: Prehybridization of filters containing DNA is carried out for 8 h to overnight at 65°C in buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 µg/ml denatured salmon sperm DNA. Filters are hybridized for 48 h at 65°C, the preferred hybridization temperature, in prehybridization mixture containing 100 µg/ml denatured salmon sperm DNA and 5-20 X 10^6 cpm of ^{32}P-labeled probe. Alternatively, the hybridization step can be performed at 65°C in the presence of SSC buffer, 1 x SSC corresponding to 0.15M NaCl and 0.05 M Na citrate. Subsequently, filter washes can be done at 37°C for 1 h in a solution containing 2 x SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA, followed by a wash in 0.1 X SSC at 50°C for 45 min. Alternatively, filter washes can be performed in a solution containing 2 x SSC and 0.1% SDS, or 0.5 x SSC and 0.1% SDS, or 0.1 x SSC and 0.1% SDS at 68°C for 15 minute intervals. Following the wash steps, the hybridized probes are detectable by autoradiography. Other conditions of high stringency which may be used are well known in the art and as cited in Sambrook et al., 1989; and Ausubel et al., 1989. These hybridization conditions are suitable for a nucleic acid molecule of about 20 nucleotides in length. There is no need to say that the hybridization conditions described above are to be adapted according to the length of the desired nucleic acid, following techniques well-known to one skilled in the art. The suitable hybridization conditions may for example be adapted according to the teachings disclosed in

the book of Hames and Higgins (*NucleicAcid Hybridization: A Practical Approach*, IRL Press, Oxford, 1985) or in Sambrook et al. (*Molecular Cloning: A Laboratory Manual*, 2nd edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989).

I. Biallelic Markers and Polynucleotides Comprising Biallelic Markers

I.A. Polynucleotides of the Present Invention

The present invention encompasses polynucleotides for use as primers and probes in the methods of the invention. These polynucleotides may consist of, consist essentially of, or comprise a contiguous span of nucleotides of a sequence from any sequence in the Sequence Listing as well as sequences which are complementary thereto ("complements thereof"). The "contiguous span" may be at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500 or 1000 nucleotides in length, to the extent that a contiguous span of these lengths is consistent with the lengths of the particular Sequence ID. It should be noted that the polynucleotides of the present invention are not limited to having the exact flanking sequences surrounding the polymorphic bases, which are enumerated in the Sequence Listing. Rather, it will be appreciated that the flanking sequences surrounding the biallelic markers, or any of the primers of probes of the invention which, are more distant from the markers, may be lengthened or shortened to any extent compatible with their intended use and the present invention specifically contemplates such sequences. It will be appreciated that the polynucleotides referred to in the Sequence Listing may be of any length compatible with their intended use. Also the flanking regions outside of the contiguous span need not be homologous to native flanking sequences which actually occur in human subjects. The addition of any nucleotide sequence, which is compatible with the nucleotides intended use is specifically contemplated. The contiguous span may optionally include the eicosanoid-related biallelic marker in said sequence. Biallelic markers generally consist of a polymorphism at one single base position. Each biallelic marker therefore corresponds to two forms of a polynucleotide sequence which, when compared with one another, present a nucleotide modification at one position. Usually, the nucleotide modification involves the substitution of one nucleotide for another. Optionally either the original or the alternative allele of the biallelic markers disclosed in Figure 4, or the first or second allele disclosed in Figure 3 and 5 may be specified as being present at the eicosanoid-related biallelic marker. Optionally, the biallelic markers may be specified as 12-214-85, 12-215-272, 12-221-163, 12-225-82, 10-234-179, 10-235-272, 10-251-342, 10-395-367, 12-730-58, 12-735-208, 12-739-22, 12-540-363, 12-550-206, 10-207-410, 10-171-254, 12-94-110, 12-834-290, 10-55-115, 12-857-122, 12-872-175, 12-882-40, 12-888-234, 12-278-353, 12-283-386, 12-44-181, 10-343-231, 10-349-216, 10-509-295, 10-511-337, 10-349-216, 10-343-231, 10-13-396, 12-570-62, 10-474-320,

10-510-173 and 10-342-301 which consist of more complex polymorphisms including insertions/deletions of at least one nucleotide. Optionally either the original or the alternative allele of these biallelic markers may be specified as being present at the eicosanoid-related biallelic marker. Preferred polynucleotides may consist of, consist essentially of, or comprise a contiguous span of nucleotides of a sequence from SEQ ID No 571-595, 600, 606, 613, 620, 622, 628, 638 and 639 as well as sequences which are complementary thereto. The "contiguous span" may be at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500 or 1000 nucleotides in length, to the extent that a contiguous span of these lengths is consistent with the lengths of the particular Sequence ID. The contiguous span may optionally comprise a biallelic marker selected from the group consisting of biallelic markers 12-214-85, 12-215-272, 12-221-163, 12-225-82, 10-234-179, 10-235-272, 10-251-342, 10-395-367, 12-730-58, 12-735-208, 12-739-22, 12-540-363, 12-550-206, 10-207-410, 10-171-254, 12-94-110, 12-834-290, 10-55-115, 12-857-122, 12-872-175, 12-882-40, 12-888-234, 12-278-353, 12-283-386, 12-44-181, 10-343-231, 10-349-216, 10-509-295, 10-511-337, 10-349-216, 10-343-231, 10-13-396, 12-570-62, 10-474-320, 10-510-173 and 10-342-301.

The invention also relates to polynucleotides that hybridize, under conditions of high or intermediate stringency, to a polynucleotide of a sequence from any sequence in the Sequence Listing as well as sequences, which are complementary thereto. Preferably such polynucleotides are at least 20, 25, 35, 40, 50, 70, 80, 100, 250, 500 or 1000 nucleotides in length, to the extent that a polynucleotide of these lengths is consistent with the lengths of the particular Sequence ID. Preferred polynucleotides comprise an eicosanoid-related biallelic marker. Optionally either the original or the alternative allele of the biallelic markers disclosed in Figure 5 may be specified as being present at the eicosanoid-related biallelic marker. Conditions of high and intermediate stringency are further described in III.C.4 "Methods of Genotyping DNA Samples for Biallelic Markers-Hybridization assay methods."

The preferred polynucleotides of the invention include the sequence ranges included in any one the sequence ranges of Figures 3, and 6 to 9 individually or in groups consisting of all the possible combinations of the ranges of included in Figures 3, and 6 to 9. The preferred polynucleotides of the invention also include fragments of at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500 or 1000 consecutive nucleotides of the sequence ranges included in any one of the sequence ranges of Figures 3, and 6 to 9 to the extent that fragments of these lengths are consistent with the lengths of the particular sequence range. The preferred polynucleotides of the invention also include fragments of at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500 or 1000 consecutive nucleotides of the sequence complementary to the sequence ranges included in any one of the sequence ranges of Figures 3, and 6 to 9 to the extent that fragments of these lengths are consistent with the lengths of the particular sequence range.

Particularly preferred polynucleotides of the invention include isolated, purified or recombinant polynucleotides comprising a contiguous span of at least 12,15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 651, wherein said contiguous span comprises at least 1, 2, 3, 4, 5 or 10 of the following nucleotide positions of SEQ ID No. 651: 1 to 2584, 4425 to 5551, 5634 to 5757, 5881 to 5995, 6100 to 6348, 6510 to 7378, 7523 to 8644, 8855 to 12253, 12341 to12853, 13024 to 13307, 13430 to 16566, 16668 to 16774, 16946 to 17062, 17555 to 20674; and the complements thereof. Other particularly preferred polynucleotides of the invention include isolated, purified or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides from a sequence of SEQ ID No. 651 and the complements thereof; wherein said contiguous span comprises at least one nucleotide positions selected from the group consisting of: a C at position 3355, a G at position 3488, a G at position 3489, and a G at position 3708 of SEQ ID No. 651.

Additional preferred polynucleotides of the invention include isolated, purified or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides from a sequence of SEQ ID No. 652, wherein said contiguous span comprises a T at position 1205 of SEQ ID No. 652 or nucleotide positions 2151 to 2157of SEQ ID No. 652; and the complements thereof.

The present invention further embodies isolated, purified, and recombinant polynucleotides which encode polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No. 653, wherein said contiguous span comprises at least one amino acid position selected from the group consisting of the following: an His residue et amino acid position 189, an His residue at amino acid position 225, a Cys residue at amino acid position 243, an Arg residue at amino acid position 261, an Asn residue at amino acid position 322, an Arg residue at amino acid position 337, a Asn residue at amino acid position 362, an Asn at amino acid position 568 and a Lys residue at amino acid position 574. The present invention further provides isolated, purified, and recombinant polynucleotides which encode polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No. 653, wherein said contiguous span comprises at least one of amino acid positions 110-131 of SEQ ID No. 653.

Particularly preferred polynucleotides of the present invention include purified, isolated or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of a sequence selected from the group consisting of SEQ ID Nos. 26-68, 614-646, and 651-652 or the complements thereof, wherein said span includes a

12-lipoxygenase-related biallelic marker. Optionally said biallelic marker is selected from the biallelic markers described in Table I (see below) and even more preferably said biallelic marker is selected from biallelic markers: 12-197-244, 12-208-35, 12-226-167, 12-206-366, 10-346-141, 10-347-111, 10-347-165, 10-347-203, 10-347-220, 10-349-97, 10-349-224, 10-341-116, 12-196-119, 12-214-129, 12-216-421, 12-219-230, and 12-223-207. Optionally either allele of the biallelic markers described above in the definition of 12-lipoxygenase-related biallelic marker is specified as being present at the 12-lipoxygenase-related biallelic marker.

Particularly preferred polynucleotides of the present invention include purified, isolated or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of a sequence of SEQ ID No. 651 and the complements thereof; wherein said contiguous span comprises a least one nucleotide positions selected from the group consisting of: a T at position 2323, a C at position 2341, an A at position 2623, an A at position 2832, a C at position 2844, an A at position 2934, an A at position 2947, a G at position 3802, a G at position 4062, a C at position 4088, a T at position 4109, a T at position 4170, an A at position 6019, a C at position 6375, a C at position 6429, an A at position 6467, a G at position 6484, an A at position 8658, a G at position 8703, an A at position 8777, a G at position 8785, a G at position 13341, an A at position 16836, an A at position 16854, and a T at position 17355 of SEQ ID No. 651.

Particularly preferred polynucleotides of the present invention include purified, isolated or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of a sequence of SEQ ID No. 652 and the complements thereof; wherein said contiguous span comprises a least one nucleotide position selected from the group consisting of: G at position 366, an A at position 605, a C at position 712, a T at position 766, an A at position 804, a G at position 821, an A at position 1004, a G at position 1049, an A at position 1123, a G at position 1131, a G at position 1491, an A at position 1742, an A at position 1760, an A at position 1941, and a T at position 2144 of SEQ ID No. 652.

Table I contains a list of preferred 12-LO-related biallelic markers. Each marker is described by indicating its Marker ID, the position of the marker in the SEQ ID and the two most common alleles.

Table I

NON-GENOMIC BIALLELIC MARKERS			
BIALLELIC MARKER ID	ALLELES	POSITION OF BIALLELIC MARKER IN SEQ ID	
		SEQ ID No.	Position
12-196-119	C/T	44	119
12-197-244	C/T	45	243
12-198-128	A/G	46	128
12-208-35	A/T	48	35
12-214-129	C/T	49	129
12-214-151	G/C	50	151
12-214-360	C/G	51	358
12-214-85	Deletion CCTAT	571	85
12-215-272	Deletion T	572	271
12-215-467	G/T	52	466
12-216-421	A/G	53	418
12-219-230	A/G	54	229
12-219-256	C/T	55	255
12-221-163	GTCCTA/T	573	163
12-221-302	A/C	57	302
12-223-179	A/G	58	179
12-223-207	C/T	59	207
12-225-541	C/T	60	540
12-225-82	Deletion T	574	82
12-226-167	C/G	61	166
12-226-458	C/T	62	455
12-229-332	G/C	63	332
12-229-351	G/C	64	351
12-230-364	C/T	65	364
12-231-100	C/T	66	99
12-231-148	C/T	67	147
12-231-266	C/T	68	265

Table I

BIALLELIC MARKERS IN GENOMIC SEQUENCE (SEQ ID No. 651)		
BIALLELIC MARKER ID	ALLELES	POSITION OF BIALLELIC MARKER IN SEQ ID
10-508-191	C/T	1128
10-508-245	C/T	1182
10-510-173	ATTTA/TTTTTT	1827
10-511-62	C/T	2048
10-511-337	Insertion of T	2323
10-512-36	G/C	2341
10-512-318	A/G	2623
10-513-250	A/G	2832
10-513-262	C /T	2844

10-513-352	A/G	2934
10-513-365	A/G	2947
12-206-81	A/G	3802
10-343-231	Deletion of C	4062
12-206-366	C/T	4088
10-343-278	C/T	4109
10-343-339	G/T	4170
10-346-23	A/G	5903
10-346-141	A/G	6019
10-346-263	G/C	6141
10-346-305	C/T	6183
10-347-74	A/G	6338
10-347-111	G/C	6375
10-347-165	C/T	6429
10-347-203	A/G	6467
10-347-220	A/G	6484
10-347-271	A/T	6534
10-347-348	A/G	6611
10-348-391	A/G	7668
10-349-47	C/T	8608
10-349-97	A/G	8658
10-349-142	G/C	8703
10-349-216	Deletion of CTG	8777
10-349-224	G/T	8785
10-349-368	C/T	8926
10-350-72	C/T	12171
10-350-332	C/T	12429
10-507-170	A/G	13341
10-507-321	A/C	13492
10-507-353	C/T	13524
10-507-364	C/T	13535
10-507-405	C/T	13576
12-220-48	G/A	15194
10-339-32	C/T	16468
10-339-124	C/T	16559
10-340-112	A/C	16836
10-340-130	A/T	16854
10-340-238	A/G	16962
10-341-116	A/G	17152
10-341-319	C/T	17355
10-342-301	Insertion of A	17623
10-342-373	C/T	17695

Table I

BIALLELIC MARKERS IN 12-LO cDNA (SEQ ID No 652)		
BIALLELIC MARKER ID	**ALLELES**	**POSITION OF BIALLELIC MARKER IN SEQ ID**
10-343-231	Deletion of C	366
10-346-141	A /G	605
10-347-111	G/C	712
10-347-165	C/T	766
10-347-203	A/G	804
10-347-220	A/G	821
10-349-142	G/C	1049
10-349-216	Deletion of CTG	1123
10-349-224	G/T	1131
10-507-170	A/G	1491
10-340-112	A/C	1742
10-340-130	A/T	1760
10-341-116	A/G	1941
10-341-319	C/T	2144

The primers of the present invention may be designed from the disclosed sequences for any method known in the art. A preferred set of primers is fashioned such that the 3' end of the contiguous span of identity with the sequences of the Sequence Listing is present at the 3' end of the primer. Such a configuration allows the 3' end of the primer to hybridize to a selected nucleic acid sequence and dramatically increases the efficiency of the primer for amplification or sequencing reactions. In a preferred set of primers the contiguous span is found in one of the sequences described in Figure 6. Allele specific primers may be designed such that a biallelic marker is at the 3' end of the contiguous span and the contiguous span is present at the 3' end of the primer. Such allele specific primers tend to selectively prime an amplification or sequencing reaction so long as they are used with a nucleic acid sample that contains one of the two alleles present at a biallelic marker. The 3' end of primers of the invention may be located within or at least 2, 4, 6, 8, 10, 12, 15, 18, 20, 25, 50, 100, 250, 500, or 1000, to the extent that this distance is consistent with the particular Sequence ID, nucleotides upstream of an eicosanoid-related biallelic marker in said sequence or at any other location which is appropriate for their intended use in sequencing, amplification or the location of novel sequences or markers. A list of preferred amplification primers is disclosed in Figure 8. A more preferred set of amplification primers is described in Figure 8 in SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652. Primers with their 3' ends located 1 nucleotide upstream of an eicosanoid-related biallelic marker have a special utility as microsequencing assays. Preferred microsequencing primers are described in Figures 7. A more preferred set of microsequencing primers is described in Figure 7 in SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652.

The probes of the present invention may be designed from the disclosed sequences for any method known in the art, particularly methods which allow for testing if a particular sequence or marker disclosed herein is present. A preferred set of probes may be designed for use in the hybridization assays of the invention in any manner known in the art such that they selectively bind to one allele of a biallelic marker, but not the other under any particular set of assay conditions. Preferred hybridization probes may consists of, consist essentially of, or comprise a contiguous span which ranges in length from 8, 10, 12, 15, 18 or 20 to 25, 35, 40, 50, 60, 70, or 80 nucleotides, or be specified as being 12, 15, 18, 20, 25, 35, 40, or 50 nucleotides in length and including an eicosanoid-related biallelic marker of said sequence. Optionally the original allele or alternative allele disclosed in Figure 4 and 5 may be specified as being present at the biallelic marker site. Optionally, said biallelic marker may be within 6, 5, 4, 3, 2, or 1 nucleotides of the center of the hybridization probe or at the center of said probe. A particularly preferred set of hybridization probes is disclosed in Figure 9 or a sequence complementary thereto.

Any of the polynucleotides of the present invention can be labeled, if desired, by incorporating a label detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive substances, fluorescent dyes or biotin. Preferably, polynucleotides are labeled at their 3' and 5' ends. A label can also be used to capture the primer, so as to facilitate the immobilization of either the primer or a primer extension product, such as amplified DNA, on a solid support. A capture label is attached to the primers or probes and can be a specific binding member which forms a binding pair with the solid's phase reagent's specific binding member (e.g. biotin and streptavidin). Therefore depending upon the type of label carried by a polynucleotide or a probe, it may be employed to capture or to detect the target DNA. Further, it will be understood that the polynucleotides, primers or probes provided herein, may, themselves, serve as the capture label. For example, in the case where a solid phase reagent's binding member is a nucleic acid sequence, it may be selected such that it binds a complementary portion of a primer or probe to thereby immobilize the primer or probe to the solid phase. In cases where a polynucleotide probe itself serves as the binding member, those skilled in the art will recognize that the probe will contain a sequence or "tail" that is not complementary to the target. In the case where a polynucleotide primer itself serves as the capture label, at least a portion of the primer will be free to hybridize with a nucleic acid on a solid phase. DNA Labeling techniques are well known to the skilled technician.

Any of the polynucleotides, primers and probes of the present invention can be conveniently immobilized on a solid support. Solid supports are known to those skilled in the art and include the walls of wells of a reaction tray, test tubes, polystyrene beads, magnetic beads, nitrocellulose strips, membranes, microparticles such as latex particles, sheep (or other animal) red blood cells, duracytes®

and others. The solid support is not critical and can be selected by one skilled in the art. Thus, latex particles, microparticles, magnetic or non-magnetic beads, membranes, plastic tubes, walls of microtiter wells, glass or silicon chips, sheep (or other suitable animal's) red blood cells and duracytes are all suitable examples. Suitable methods for immobilizing nucleic acids on solid phases include ionic, hydrophobic, covalent interactions and the like. A solid support, as used herein, refers to any material which is insoluble, or can be made insoluble by a subsequent reaction. The solid support can be chosen for its intrinsic ability to attract and immobilize the capture reagent. Alternatively, the solid phase can retain an additional receptor which has the ability to attract and immobilize the capture reagent. The additional receptor can include a charged substance that is oppositely charged with respect to the capture reagent itself or to a charged substance conjugated to the capture reagent. As yet another alternative, the receptor molecule can be any specific binding member which is immobilized upon (attached to) the solid support and which has the ability to immobilize the capture reagent through a specific binding reaction. The receptor molecule enables the indirect binding of the capture reagent to a solid support material before the performance of the assay or during the performance of the assay. The solid phase thus can be a plastic, derivatized plastic, magnetic or non-magnetic metal, glass or silicon surface of a test tube, microtiter well, sheet, bead, microparticle, chip, sheep (or other suitable animal's) red blood cells, duracytes® and other configurations known to those of ordinary skill in the art. The polynucleotides of the invention can be attached to or immobilized on a solid support individually or in groups of at least 2, 5, 8, 10, 12, 15, 20, or 25 distinct polynucleotides of the inventions to a single solid support. In addition, polynucleotides other than those of the invention may be attached to the same solid support as one or more polynucleotides of the invention.

Any polynucleotide provided herein may be attached in overlapping areas or at random locations on the solid support. Alternatively the polynucleotides of the invention may be attached in an ordered array wherein each polynucleotide is attached to a distinct region of the solid support which does not overlap with the attachment site of any other polynucleotide. Preferably, such an ordered array of polynucleotides is designed to be "addressable" where the distinct locations are recorded and can be accessed as part of an assay procedure. Addressable polynucleotide arrays typically comprise a plurality of different oligonucleotide probes that are coupled to a surface of a substrate in different known locations. The knowledge of the precise location of each polynucleotides location makes these "addressable" arrays particularly useful in hybridization assays. Any addressable array technology known in the art can be employed with the polynucleotides of the invention. One particular embodiment of these polynucleotide arrays is known as the Genechips™, and has been generally described in US Patent 5,143,854; PCT publications WO 90/15070 and 92/10092. These arrays may generally be produced using mechanical synthesis methods or light directed synthesis methods, which

incorporate a combination of photolithographic methods and solid phase oligonucleotide synthesis (Fodor et al., *Science*, 251:767-777, 1991). The immobilization of arrays of oligonucleotides on solid supports has been rendered possible by the development of a technology generally identified as "Very Large Scale Immobilized Polymer Synthesis" (VLSIPS™) in which, typically, probes are immobilized in a high density array on a solid surface of a chip. Examples of VLSIPS™ technologies are provided in US Patents 5,143,854 and 5,412,087 and in PCT Publications WO 90/15070, WO 92/10092 and WO 95/11995, which describe methods for forming oligonucleotide arrays through techniques such as light-directed synthesis techniques. In designing strategies aimed at providing arrays of nucleotides immobilized on solid supports, further presentation strategies were developed to order and display the oligonucleotide arrays on the chips in an attempt to maximize hybridization patterns and sequence information. Examples of such presentation strategies are disclosed in PCT Publications WO 94/12305, WO 94/11530, WO 97/29212 and WO 97/31256.

Oligonucleotide arrays may comprise at least one of the sequences selected from the group consisting of SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652, and the sequences complementary thereto or a fragment thereof of at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500 or 1000 consecutive nucleotides, to the extent that fragments of these lengths is consistent with the lengths of the particular Sequence ID, for determining whether a sample contains one or more alleles of the biallelic markers of the present invention. Oligonucleotide arrays may also comprise at least one of the sequences selected from the group consisting of SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652, and the sequences complementary thereto or a fragment thereof of at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500 or 1000 consecutive nucleotides, to the extent that fragments of these lengths is consistent with the lengths of the particular Sequence ID, for amplifying one or more alleles of the biallelic markers of Figure 2. In other embodiments, arrays may also comprise at least one of the sequences selected from the group consisting of SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652; and the sequences complementary thereto or a fragment thereof of at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500 or 1000 consecutive nucleotides, to the extent that fragments of these lengths is consistent with the lengths of the particular Sequence ID, for conducting microsequencing analyses to determine whether a sample contains one or more alleles of the biallelic markers of the invention. In still further embodiments, the oligonucleotide array may comprise at least one of the sequences selecting from the group consisting of SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652; and the sequences

complementary thereto or a fragment thereof of at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500 or 1000 nucleotides in length, to the extent that fragments of these lengths is consistent with the lengths of the particular Sequence ID, for determining whether a sample contains one or more alleles of the biallelic markers of the present invention. In still further embodiments, the oligonucleotide array may comprise at least one of the novel sequences listed in the fifth column of Figure 3 or the sequences complementary thereto or a fragment comprising at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500 or 1000 consecutive nucleotides thereof to the extent that fragments of these lengths are consistent with the lengths of the particular novel sequences.

The present invention also encompasses diagnostic kits comprising one or more polynucleotides of the invention, optionally with a portion or all of the necessary reagents and instructions for genotyping a test subject by determining the identity of a nucleotide at an eicosanoid-related biallelic marker. The polynucleotides of a kit may optionally be attached to a solid support, or be part of an array or addressable array of polynucleotides. The kit may provide for the determination of the identity of the nucleotide at a marker position by any method known in the art including, but not limited to, a sequencing assay method, a microsequencing assay method, a hybridization assay method, an allele specific amplification method, or a mismatch detection assay based on polymerases and/or ligases. Optionally such a kit may include instructions for scoring the results of the determination with respect to the test subjects' risk of contracting a diseases involving arachidonic acid metabolism, or likely response to an agent acting on arachidonic acid metabolism, or chances of suffering from side effects to an agent acting on arachidonic acid metabolism. Preferably such a kit may include instructions for scoring the results of the determination with respect to the subjects risk of developing hepatotoxicity upon treatment with the anti-asthmatic drug zileuton.

It should be noted that in the accompanying Sequence Listing, all instances of the symbol "n" in the nucleic acid sequences mean that the nucleotide can be adenine, guanine, cytosine or thymine.

I.B. Genomic Sequences of the 12-LO Gene and Biallelic Markers

The present invention encompasses the genomic sequence of the 12-LO gene of SEQ ID No. 651. The 12-LO genomic sequences comprise exons and introns. Particularly preferred genomic sequences of the 12-LO gene include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12,15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No. 651, wherein said contiguous span comprises at least 1 one of the following nucleotide positions of SEQ ID No. 651: 1 to 2584, 4425 to 5551, 5634 to 5757, 5881 to 5995, 6100 to 6348, 6510 to 7378, 7523 to 8644, 8855 to 12253, 12341 to12853, 13024 to 13307, 13430 to 16566, 16668 to 16774, 16946 to 17062, 17555 to 20674; and the complements thereof. The nucleic acids

defining the 12-LO intronic polynucleotides may be used as oligonucleotide primers or probes in order to detect the presence of a copy of the 12-LO gene in a test sample, or alternatively in order to amplify a target nucleotide sequence within the12-LO sequences. Other particularly preferred genomic sequences of the invention include isolated, purified or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides from a sequence of SEQ ID No. 651 and the complements thereof; wherein said contiguous span comprises at least one nucleotide positions selected from the group consisting of: a C at position 3355, a G at position 3488, a G at position 3489, and a G at position 3708 of SEQ ID No. 651.

The present invention further provides 12-lipoxygenase intron and exon polynucleotide sequences including biallelic markers. Particularly preferred polynucleotides of the present invention include purified, isolated or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of a sequence of SEQ ID No. 651 or the complements thereof, wherein said span includes a 12-lipoxygenase-related biallelic marker. Optionally said biallelic marker is selected from the biallelic markers described in Table I and even more preferably said biallelic marker is selected from biallelic markers: 12-197-244, 12-208-35, 12-226-167, 12-206-366, 10-346-141, 10-347-111, 10-347-165, 10-347-203, 10-347-220, 10-349-97, 10-349-224, 10-341-116, 12-196-119, 12-214-129, 12-216-421, 12-219-230, and 12-223-207. Particularly preferred genomic sequences of the present invention include purified, isolated or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of a sequence of SEQ ID No. 651 and the complements thereof; wherein said contiguous span comprises a least one nucleotide positions selected from the group consisting of: a T at position 2323, a C at position 2341, an A at position 2623, an A at position 2832, a C at position 2844, an A at position 2934, an A at position 2947, a G at position 3802, a G at position 4062, a C at position 4088, a T at position 4109, a T at position 4170, an A at position 6019, a C at position 6375, a C at position 6429, an A at position 6467, a G at position 6484, an A at position 8658, a G at position 8703, an A at position 8777, a G at position 8785, a G at position 13341, an A at position 16836, an A at position 16854, and a T at position 17355 of SEQ ID No. 651.

The genomic sequence of the 12-LO gene contains regulatory sequences both in the non-coding 5'- flanking region and in the non-coding 3'- flanking region that border the 12-LO transcribed region containing the 14 exons of this gene. 5'-regulatory sequences of the 12-LO gene comprise the polynucleotide sequences located between the nucleotide in position 1 and the nucleotide in position 3124 of the nucleotide sequence of SEQ ID No. 651, more preferably between positions 1 and 2195 of SEQ ID No. 651. 3'-regulatory sequences of the 12-LO gene comprise the polynucleotide sequences

located between the nucleotide in position 17555 and the nucleotide in position 20674 of the nucleotide sequence of SEQ ID No. 651.

The promoter activity of the regulatory regions contained in the 12-LO gene of polynucleotide sequence of SEQ ID No. 651 can be assessed by any known method. Methods for identifying the polynucleotide fragments of SEQ ID No. 651 involved in the regulation of the expression of the 12-LO gene are well-known to those skilled in the art (see Sambrook et al., *Molecular Cloning A Laboratory Manual*, 2nd edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989). An example of a typical method, that can be used, involves a recombinant vector carrying a reporter gene and genomic sequences from the 12-LO genomic sequence of SEQ ID No. 651. Briefly, the expression of the reporter gene (for example beta galactosidase or chloramphenicol acetyl transferase) is detected when placed under the control of a biologically active polynucleotide fragment. Genomic sequences located upstream of the first exon of the 12-LO gene may be cloned into any suitable promoter reporter vector, such as the pSEAP-Basic, pSEAP-Enhancer, pβgal-Basic, pβgal-Enhancer, or pEGFP-1 Promoter Reporter vectors available from Clontech, or pGL2-basic or pGL3-basic promoterless luciferase reporter gene vector from Promega. Each of these promoter reporter vectors include multiple cloning sites positioned upstream of a reporter gene encoding a readily assayable protein such as secreted alkaline phosphatase, luciferase, beta galactosidase, or green fluorescent protein. The sequences upstream the first 12-LO exon are inserted into the cloning sites upstream of the reporter gene in both orientations and introduced into an appropriate host cell. The level of reporter protein is assayed and compared to the level obtained with a vector lacking an insert in the cloning site. The presence of an elevated expression level in the vector containing the insert with respect to the control vector indicates the presence of a promoter in the insert.

Promoter sequences within the 5' non-coding regions of the 12-LO gene may be further defined by constructing nested 5' and/or 3' deletions using conventional techniques such as Exonuclease III or appropriate restriction endonuclease digestion. The resulting deletion fragments can be inserted into the promoter reporter vector to determine whether the deletion has reduced or obliterated promoter activity, such as described, for example, by Coles et al. (*Hum. Mol. Genet.*, 7:791-800, 1998). In this way, the boundaries of the promoters may be defined. If desired, potential individual regulatory sites within the promoter may be identified using site directed mutagenesis or linker scanning to obliterate potential transcription factor binding sites within the promoter individually or in combination. The effects of these mutations on transcription levels may be determined by inserting the mutations into cloning sites in promoter reporter vectors. This type of assays are well known to those skilled in the art and are further described in WO 97/17359, US 5 374 544, EP 582 796, US 5 698 389, US 5 643 746, US 5 502 176, and US 5 266 488.

The activity and the specificity of the promoter of the 12-LO gene can further be assessed by monitoring the expression level of a detectable polynucleotide operably linked to the 12-LO promoter in different types of cells and tissues. The detectable polynucleotide may be either a polynucleotide that specifically hybridizes with a predefined oligonucleotide probe, or a polynucleotide encoding a detectable protein, including a 12-LO polypeptide or a fragment or a variant thereof. This type of assay is well known to those skilled in the art and is described in US 5 502 176, and US 5 266 488.

Polynucleotides carrying the regulatory elements located both at the 5' end and at the 3' end of the 12-LO coding region may be advantageously used to control the transcriptional and translational activity of a heterologous polynucleotide of interest, said polynucleotide being heterologous as regards to the 12-LO regulatory region.

Thus, the present invention also concerns a purified, isolated, and recombinant nucleic acid comprising a polynucleotide which, is selected from the group consisting of, the polynucleotide sequences located between the nucleotide in position 1 and the nucleotide in position 3124 of the nucleotide sequence of SEQ ID No. 651, more preferably between positions 1 and 2195 of SEQ ID No. 651 and the polynucleotide sequences located between the nucleotide in position 17555 and the nucleotide in position 20674 of SEQ ID No. 651; or a sequence complementary thereto or a biologically active fragment thereof.

A "biologically active" fragment of SEQ ID No. 651 according to the present invention is a polynucleotide comprising or alternatively consisting of a fragment of said polynucleotide which is functional as a regulatory region for expressing a recombinant polypeptide or a recombinant polynucleotide in a recombinant cell host.

For the purpose of the invention, a nucleic acid or polynucleotide is "functional" as a regulatory region for expressing a recombinant polypeptide or a recombinant polynucleotide if said regulatory polynucleotide contains nucleotide sequences which contain transcriptional and translational regulatory information, and such sequences are "operably linked" to nucleotide sequences which encode the desired polypeptide or the desired polynucleotide.

The regulatory polynucleotides according to the invention may be advantageously part of a recombinant expression vector that may be used to express a coding sequence in a desired host cell or host organism.

A further object of the invention consists of an isolated polynucleotide comprising:

a) a nucleic acid comprising a regulatory nucleotide sequence selected from the group consisting of a nucleotide sequence comprising a polynucleotide of SEQ ID No. 651;

b) a polynucleotide encoding a desired polypeptide or a nucleic acid of interest, operably linked to the nucleic acid defined in (a) above.

The polypeptide encoded by the nucleic acid described above may be of various nature or origin, encompassing proteins of prokaryotic or eukaryotic origin. Among the polypeptides expressed under the control of a 12-LO regulatory region, there may be cited bacterial, fungal or viral antigens. Also encompassed are eukaryotic proteins such as intracellular proteins, for example "house keeping" proteins, membrane-bound proteins, for example receptors, and secreted proteins, for example cytokines. In a specific embodiment, the desired polypeptide may be the 12-LO protein, especially the protein of the amino acid sequence of SEQ ID No. 653 and 654.

The desired nucleic acids encoded by the above described polynucleotide, usually a RNA molecule, may be complementary to a desired coding polynucleotide, for example to the 12-LO coding sequence, and thus useful as an antisense polynucleotide.

Such a polynucleotide may be included in a recombinant expression vector in order to express the desired polypeptide or the desired nucleic acid in host cell or in a host organism.

C. cDNA Sequences of the 12-LO Gene and Biallelic Markers

The present invention provides a 12-lipoxygenase cDNA of SEQ ID No. 652. The Open Reading Frame encoding the 12-LO protein spans from the nucleotide in position 40 to the nucleotide in position 2028 of the polynucleotide sequence of SEQ ID No. 652. The cDNA of SEQ ID No. 652 also includes a 5'-UTR region (1-40) and a 3'-UTR (2028-2343) region.

Additional preferred cDNA polynucleotides of the invention include isolated, purified or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides from a sequence of SEQ ID No. 652 and the complements thereof. Additional preferred polynucleotides include isolated, purified or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides from a sequence of SEQ ID No. 652, wherein said contiguous span comprises a T at position 1205 of SEQ ID No. 652 or nucleotide positions 2151 to 2157of SEQ ID No. 652; and the complements thereof.

Preferred cDNA fragments comprise a biallelic marker selected from the group consisting of 10-343-231, 10-346-141, 10-347-111, 10-347-165, 10-347-203, 10-347-220, 10-349-97, 10-349-142, 10-349-216, 10-349-224, 10-507-170, 10-340-112, 10-340-130, 10-341-116 and 10-341-319. Some biallelic polymorphisms represent silent nucleotide substitutions but biallelic markers 10-346-141, 10-347-111, 10-347-165, 10-347-220, 10-349-97, 10-349-142, 10-349-216, 10-340-112, 10-340-130 are associated with amino acid changes in the corresponding 12-lipoxygenase polypeptide. One allele of biallelic marker 10-343-231 (polymorphic deletion of a C nucleotide at position 366 of SEQ ID No. 652) causes a frame shift in the open reading frame of the 12-LO cDNA of SEQ ID No. 652 resulting in

the novel polypeptide of SEQ ID No. 653. 12-LO polypeptides of SEQ ID Nos. 653 and 654 of the present invention are further described below.

Other preferred cDNA fragments comprise a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides from a sequence of SEQ ID No. 652, wherein said contiguous span comprises a T at position 1205 of SEQ ID No. 652; and the complements thereof. 12-LO cDNA fragments comprise a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides from a sequence of SEQ ID No. 652, wherein said contiguous span comprises a T at position 1205 of SEQ ID No. 652 encode novel 12-LO polypeptides of SEQ ID No. 653 comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No. 653, wherein said contiguous span comprises a Leu residue at amino acid position 389 of SEQ ID No. 653.

The polynucleotide disclosed above that contains the coding sequence of the 12-LO gene of the invention may be expressed in a desired host cell or a desired host organism, when this polynucleotide is placed under the control of suitable expression signals. The expression signals may be either the expression signals contained in the regulatory regions in the 12-LO gene of the invention or may be exogenous regulatory nucleic sequences. Such a polynucleotide, when placed under the suitable expression signals, may also be inserted in a vector for its expression.

Another preferred cDNA fragment comprises the 5'-UTR (5'regulatory sequence) region beginning at position 1 and ending at position 39 of SEQ ID No. 652. Another preferred cDNA fragment comprises the 3'-UTR (3'regulatory sequence) region beginning at position 2029 and ending at position 2343 of SEQ ID No. 652. Preferably said 3'-UTR region comprises biallelic marker 10-341-319 or nucleotide positions 2151 to 2157of SEQ ID No. 652.

I.D. Polynucleotide Constructs, Recombinant Vectors, Host Cells and Transgenic Animals

The terms "polynucleotide construct" and "recombinant polynucleotide" are used interchangeably herein to refer to linear or circular, purified or isolated polynucleotides that have been artificially designed and which comprise at least two nucleotide sequences that are not found as contiguous nucleotide sequences in their initial natural environment.

i. Polynucleotide constructs

1. DNA constructs for expressing the 12-LO gene in recombinant host cells and in transgenic animals

In order to study the physiological and phenotype consequences of a lack of synthesis of the 12-LO protein, both at the cellular level and at the multicellular organism level, in particular as regards to disorders related to abnormal cell proliferation, notably cancers, the invention also encompasses DNA constructs and recombinant vectors enabling a conditional expression of a specific allele of the 12-LO genomic sequence or cDNA

A first preferred DNA construct is based on the tetracycline resistance operon *tet* from *E. coli* transposon Tn110 for controlling the12-LO gene expression, such as described by Gossen et al. (*Science*, 268:1766-1769, 1995). Such a DNA construct contains seven *tet* operator sequences from Tn10 (*tet*op) that are fused to either a minimal promoter or a 5'-regulatory sequence of the 12-LO gene, said minimal promoter or said 12-LO regulatory sequence being operably linked to a polynucleotide of interest that codes either for a sense or an antisense oligonucleotide or for a polypeptide, including a 12-LO polypeptide or a peptide fragment thereof. This DNA construct is functional as a conditional expression system for the nucleotide sequence of interest when the same cell also comprises a nucleotide sequence coding for either the wild type (tTA) or the mutant (rTA) repressor fused to the activating domain of viral protein VP16 of herpes simplex virus, placed under the control of a promoter, such as the HCMVIE1 enhancer/promoter or the MMTV-LTR. Indeed, a preferred DNA construct of the invention will comprise both the polynucleotide containing the *tet* operator sequences and the polynucleotide containing a sequence coding for the tTA or the rTA repressor. In the specific embodiment wherein the conditional expression DNA construct contains the sequence encoding the mutant tetracycline repressor rTA, the expression of the polynucleotide of interest is silent in the absence of tetracycline and induced in its presence.

2. DNA constructs allowing homologous recombination: replacement vectors

A second preferred DNA construct will comprise, from 5'-end to 3'-end: (a) a first nucleotide sequence that is comprised in the 12-LO genomic sequence; (b) a nucleotide sequence comprising a positive selection marker, such as the marker for neomycine resistance (*neo*); and (c) a second nucleotide sequence that is comprised in the 12-LO genomic sequence, and is located on the genome downstream the first 12-LO nucleotide sequence (a).

In a preferred embodiment, this DNA construct also comprises a negative selection marker located upstream the nucleotide sequence (a) or downstream the nucleotide sequence (c). Preferably, the negative selection marker consists of the thymidine kinase (*tk*) gene (Thomas et al., *Cell*, 44:419-428, 1986), the hygromycine beta gene (Te Riele et al., *Nature*, 348:649-651, 1990), the *hprt* gene (Van der Lugt et al., *Gene*, 105:263-267, 1991; Reid et al., *Proc. Natl. Acad. Sci. USA*, 87:4299-4303, 1990) or the Diphteria toxin A fragment (*Dt-A*) gene (Nada et al., *Cell*, 73:1125-1135, 1993; Yagi et al., *Proc.*

Natl; Acad. Sci. USA, 87:9918-9922, 1990). Preferably, the positive selection marker is located within a 12-LO exon sequence so as to interrupt the sequence encoding a 12-LO protein.

These replacement vectors are further described by Mansour et al. (*Nature*, 336:348-352, 1988) and Koller et al. (*Ann. Rev. Immunol.*, 10:705-730, 1992).

The first and second nucleotide sequences (a) and (c) may be indifferently located within a 12-LO regulatory sequence, an intronic sequence, an exon sequence or a sequence containing both regulatory and/or intronic and/or exon sequences. The size of the nucleotide sequences (a) and (c) is ranging from 1 to 50 kb, preferably from 1 to 10 kb, more preferably from 2 to 6 kb and most preferably from 2 to 4 kb.

3. DNA constructs allowing homologous recombination: Cre-loxP system

These new DNA constructs make use of the site specific recombination system of the P1 phage. The P1 phage possesses a recombinase called Cre which, interacts specifically with a 34 base pairs *lox*P site. The *lox*P site is composed of two palindromic sequences of 13 bp separated by a 8 bp conserved sequence (Hoess et al., *Nucleic Acids Res.*, 14:2287-2300, 1986). The recombination by the Cre enzyme between two *lox*P sites having an identical orientation leads to the deletion of the DNA fragment.

The Cre-*lox*P system used in combination with a homologous recombination technique was first described by Gu et al. (*Cell*, 73:1155-1164, 1993). Briefly, a nucleotide sequence of interest to be inserted in a targeted location of the genome harbors at least two *lox*P sites in the same orientation and located at the respective ends of a nucleotide sequence to be excised from the recombinant genome. The excision event requires the presence of the recombinase (Cre) enzyme within the nucleus of the recombinant cell host. The recombinase enzyme may be brought at the desired time either by (a) incubating the recombinant cell hosts in a culture medium containing this enzyme, by injecting the Cre enzyme directly into the desired cell, such as described by Araki et al. (*Proc. Natl; Acad. Sci. USA*, 92: 160-164, 1995), or by lipofection of the enzyme into the cells, such as described by Baubonis et al. (*Nucleic Acids Res.*, 21:2025-2029, 1993); (b) transfecting the cell host with a vector comprising the *Cre* coding sequence operably linked to a promoter functional in the recombinant cell host, which promoter being optionally inducible, said vector being introduced in the recombinant cell host, such as described by Gu et al. (*Cell*, 73:1155-1164, 1993) and Sauer et al. (*Proc. Natl; Acad. Sci. USA*, 85:5166-5170, 1988); (c) introducing in the genome of the cell host a polynucleotide comprising the *Cre* coding sequence operably linked to a promoter functional in the recombinant cell host, which promoter is optionally inducible, and said polynucleotide being inserted in the genome of the cell host

either by a random insertion event or an homologous recombination event, such as described by Gu et al. (*Science*, 265:103-106, 1994).

In the specific embodiment wherein the vector containing the sequence to be inserted in the 12-LO gene by homologous recombination is constructed in such a way that selectable markers are flanked by *lox*P sites of the same orientation, it is possible, by treatment by the Cre enzyme, to eliminate the selectable markers while leaving the 12-LO sequences of interest that have been inserted by an homologous recombination event. Again, two selectable markers are needed: a positive selection marker to select for the recombination event and a negative selection marker to select for the homologous recombination event. Vectors and methods using the Cre-*lox*P system are further described by Zou et al. (*Curr. Biol.*, 4:1099-1103, 1994).

Thus, a third preferred DNA construct of the invention comprises, from 5'-end to 3'-end: (a) a first nucleotide sequence that is comprised in the 12-LO genomic sequence; (b) a nucleotide sequence comprising a polynucleotide encoding a positive selection marker, said nucleotide sequence comprising additionally two sequences defining a site recognized by a recombinase, such as a *lox*P site, the two sites being placed in the same orientation; and (c) a second nucleotide sequence that is comprised in the 12-LO genomic sequence, and is located on the genome downstream of the first 12-LO nucleotide sequence (a).

The sequences defining a site recognized by a recombinase, such as a *lox*P site, are preferably located within the nucleotide sequence (b) at suitable locations bordering the nucleotide sequence for which the conditional excision is sought. In one specific embodiment, two *lox*P sites are located at each side of the positive selection marker sequence, in order to allow its excision at a desired time after the occurrence of the homologous recombination event.

In a preferred embodiment of a method using the third DNA construct described above, the excision of the polynucleotide fragment bordered by the two sites recognized by a recombinase, preferably two *lox*P sites, is performed at a desired time, due to the presence within the genome of the recombinant cell host of a sequence encoding the Cre enzyme operably linked to a promoter sequence, preferably an inducible promoter, more preferably a tissue-specific promoter sequence and most preferably a promoter sequence which is both inducible and tissue-specific, such as described by Gu et al. (*Science*, 265:103-106, 1994).

The presence of the Cre enzyme within the genome of the recombinant cell host may result of the breeding of two transgenic animals, the first transgenic animal bearing the 12-LO-derived sequence of interest containing the *lox*P sites as described above and the second transgenic animal bearing the *Cre* coding sequence operably linked to a suitable promoter sequence, such as described by Gu et al. (*Science*, 265:103-106, 1994).

Spatio-temporal control of the Cre enzyme expression may also be achieved with an adenovirus based vector that contains the Cre gene thus allowing infection of cells, or *in vivo* infection of organs, for delivery of the Cre enzyme, such as described by Anton et al. (*J. Virol.*, 69:4600-4606, 1995) and Kanégae et al. (*Nucleic Acids Res.*, 23:3816-3821, 1995).

The DNA constructs described above may be used to introduce a desired nucleotide sequence of the invention, preferably a 12-LO genomic sequence or a 12-LO cDNA sequence, and most preferably an altered copy of a 12-LO genomic or cDNA sequence, within a predetermined location of the targeted genome, leading either to the generation of an altered copy of a targeted gene (knock-out homologous recombination) or to the replacement of a copy of the targeted gene by another copy sufficiently homologous to allow an homologous recombination event to occur (knock-in homologous recombination).

ii. Recombinant vectors

The term "vector" is used herein to designate either a circular or a linear DNA or RNA molecule, which is either double-stranded or single-stranded, and which comprise at least one polynucleotide of interest that is sought to be transferred in a cell host or in a unicellular or multicellular host organism.

The present invention encompasses a family of recombinant vectors that comprise a regulatory polynucleotide derived from the 12-LO genomic sequence, or a coding polynucleotide from the 12-LO genomic sequence. Consequently, the present invention further deals with a recombinant vector comprising either a regulatory polynucleotide comprised in the nucleic acid of SEQ ID No. 651 or a polynucleotide comprising the 12-LO coding sequence or both.

In a first preferred embodiment, a recombinant vector of the invention is used to amplify the inserted polynucleotide derived from a 12-LO genomic sequence selected from the group consisting of the nucleic acids of SEQ ID No. 651 or a 12-LO cDNA, for example the cDNA of SEQ ID No. 652 in a suitable host cell, this polynucleotide being amplified each time the recombinant vector replicates. Generally, a recombinant vector of the invention may comprise any of the polynucleotides described herein, including regulatory sequences and coding sequences, as well as any 12-LO primer or probe as defined above.

In a second preferred embodiment, recombinant vectors of the invention consist of expression vectors comprising either a regulatory polynucleotide or a coding nucleic acid of the invention, or both. Within certain embodiments, expression vectors are employed to express the 12-LO polypeptide which can be then purified and, for example be used in ligand screening assays or as an immunogen in order to raise specific antibodies directed against the 12-LO protein. In other embodiments, the expression

vectors are used for constructing transgenic animals and also for gene therapy. Expression requires that appropriate signals are provided in the vectors, said signals including various regulatory elements, such as enhancers/promoters from both viral and mammalian sources that drive expression of the genes of interest in host cells. Dominant drug selection markers for establishing permanent, stable cell clones expressing the products are generally included in the expression vectors of the invention, as they are elements that link expression of the drug selection markers to expression of the polypeptide.

More particularly, the present invention relates to expression vectors which include nucleic acids encoding a 12-LO protein, preferably the 12-LO protein of the amino acid sequence of SEQ ID No. 653, under the control of a regulatory sequence selected among the12-LO regulatory polynucleotides of SEQ ID Nos. 651 and 652, or alternatively under the control of an exogenous regulatory sequence.

Consequently, preferred expression vectors of the invention are selected from the group consisting of: (a) the 12-LO regulatory sequence comprised therein drives the expression of a coding polynucleotide operably linked thereto; (b) the 12-LO coding sequence is operably linked to regulation sequences allowing its expression in a suitable cell host and/or host organism.

Additionally, the recombinant expression vector described above may also comprise a nucleic acid comprising a 5'-regulatory polynucleotide, preferably a 5'-regulatory polynucleotide of the 12-LO gene. Additionally, the recombinant expression vector described above may also comprise a nucleic acid comprising a 3'-regulatory polynucleotide, preferably a 3'-regulatory polynucleotide of the 12-LO gene. The 12-LO 3'-regulatory polynucleotide may also comprise the 3'-UTR sequence contained in the nucleotide sequence of SEQ ID No. 652. The 5'-regulatory polynucleotide may also include the 5'-UTR sequence of the 12-LO cDNA, or a biologically active fragment or variant thereof. The invention also pertains to a recombinant expression vector useful for the expression of the 12-LO coding sequence, wherein said vector comprises a nucleic acid of SEQ ID No. 652.

The invention also relates to a recombinant expression vector comprising a nucleic acid comprising the nucleotide sequence beginning at the nucleotide in position 40 and ending in position 2028 of the polynucleotide of SEQ ID No. 652.

Some of the elements which can be found in the vectors of the present invention are described in further detail in the following sections.

1. General features of the expression vectors of the invention

A recombinant vector according to the invention comprises, but is not limited to, a YAC (Yeast Artificial Chromosome), a BAC (Bacterial Artificial Chromosome), a phage, a phagemid, a cosmid, a plasmid or even a linear DNA molecule which may consist of a chromosomal, non-chromosomal, semi-

synthetic and synthetic DNA. Such a recombinant vector can comprise a transcriptional unit comprising an assembly of :

(1) a genetic element or elements having a regulatory role in gene expression, for example promoters or enhancers. Enhancers are cis-acting elements of DNA, usually from about 10 to 300 bp in length that act on the promoter to increase the transcription.

(2) a structural or coding sequence which is transcribed into mRNA and eventually translated into a polypeptide, said structural or coding sequence being operably linked to the regulatory elements described in (1); and

(3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, when a recombinant protein is expressed without a leader or transport sequence, it may include a N-terminal residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product. Generally, recombinant expression vectors will include origins of replication, selectable markers permitting transformation of the host cell, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably a leader sequence capable of directing secretion of the translated protein into the periplasmic space or the extracellular medium. In a specific embodiment wherein the vector is adapted for transfecting and expressing desired sequences in mammalian host cells, preferred vectors will comprise an origin of replication in the desired host, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5'-flanking non-transcribed sequences. DNA sequences derived from the SV40 viral genome, for example SV40 origin, early promoter, enhancer, splice and polyadenylation sites may be used to provide the required non-transcribed genetic elements.

The *in vivo* expression of a 12-LO polypeptide of SEQ ID Nos. 653 and 654 may be useful in order to correct a genetic defect related to the expression of the native gene in a host organism or to the production of a biologically inactive 12-LO protein.

Consequently, the present invention also deals with recombinant expression vectors mainly designed for the *in vivo* production of the 12-LO polypeptide of SEQ ID Nos. 653-654 or fragments or variants thereof by the introduction of the appropriate genetic material in the organism of the patient to be treated. This genetic material may be introduced *in vitro* in a cell that has been previously extracted from the organism, the modified cell being subsequently reintroduced in the said organism, directly *in vivo* into the appropriate tissue.

2. Regulatory elements

The suitable promoter regions used in the expression vectors according to the present invention are chosen taking into account the cell host in which the heterologous gene has to be expressed. The particular promoter employed to control the expression of a nucleic acid sequence of interest is not believed to be important, so long as it is capable of directing the expression of the nucleic acid in the targeted cell. Thus, where a human cell is targeted, it is preferable to position the nucleic acid coding region adjacent to and under the control of a promoter that is capable of being expressed in a human cell, such as, for example, a human or a viral promoter.

A suitable promoter may be heterologous with respect to the nucleic acid for which it controls the expression or alternatively can be endogenous to the native polynucleotide containing the coding sequence to be expressed. Additionally, the promoter is generally heterologous with respect to the recombinant vector sequences within which the construct promoter/coding sequence has been inserted.

Promoter regions can be selected from any desired gene using, for example, CAT (chloramphenicol transferase) vectors and more preferably pKK232-8 and pCM7 vectors.

Preferred bacterial promoters are the LacI, LacZ, the T3 or T7 bacteriophage RNA polymerase promoters, the gpt, lambda PR, PL and trp promoters (EP 0036776), the polyhedrin promoter, or the p10 protein promoter from baculovirus (Kit Novagen) (Smith et al., *Mol. Cell.Biol.*3:2156-2165, 1983; O'Reilly et al., *Baculovirus Expression Vectors: A Laboratory Manual.*, W.H. Freeman andCo., New York, 1992), the lambda PR promoter or also the trc promoter.

Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-L. Selection of a convenient vector and promoter is well within the level of ordinary skill in the art.

The choice of a promoter is well within the ability of a person skilled in the field of genetic engineering. For example, one may refer to the book of Sambrook et al. (*Molecular Cloning: A Laboratory Manual*, 2nd edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989).

Where a cDNA insert is employed, one will typically desire to include a polyadenylation signal to effect proper polyadenylation of the gene transcript. The nature of the polyadenylation signal is not believed to be crucial to the successful practice of the invention, and any such sequence may be employed such as human growth hormone and SV40 polyadenylation signals. Also contemplated as an element of the expression cassette is a terminator. These elements can serve to enhance message levels and to minimize read through from the cassette into other sequences.

The vector containing the appropriate DNA sequence as described above, more preferably 12-LO gene regulatory polynucleotide, a polynucleotide encoding the 12-LO polypeptide of SEQ ID Nos.

653 and 654 or both of them, can be utilized to transform an appropriate host to allow the expression of the desired polypeptide or polynucleotide.

3. Selectable markers

Such markers would confer an identifiable change to the cell permitting easy identification of cells containing the expression construct. The selectable marker genes for selection of transformed host cells are preferably dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, TRP1 for *S. cerevisiae* or tetracycline, rifampicin or ampicillin resistance in *E. coli*, or levan saccharase for mycobacteria, this latter marker being a negative selection marker.

4. Preferred vectors

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and a bacterial origin of replication derived from commercially available plasmids comprising genetic elements of pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia, Uppsala, Sweden), and GEM1 (Promega Biotec, Madison, WI, USA). Large numbers of other suitable vectors are known to those of skill in the art, and commercially available, such as the following bacterial vectors : pQE70, pQE60, pQE-9 (Qiagen), pbs, pD10, phagescript, psiX174, pbluescript SK, pbsks, pNH8A, pNH16A, pNH18A, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia); pWLNEO, pSV2CAT, pOG44, pXT1, pSG (Stratagene); pSVK3, pBPV, pMSG, pSVL (Pharmacia); pQE-30 (QIAexpress).

The P1 bacteriophage vector may contain large inserts ranging from about 80 to about 100 kb. The construction of P1 bacteriophage vectors such as p158 or p158/neo8 are described by Sternberg (*Mamm. Genome*, 5:397-404, 1994). Recombinant P1 clones comprising 12-LO nucleotide sequences may be designed for inserting large polynucleotides of more than 40 kb (Linton et al., *J. Clin. Invest.*, 92:3029-3037, 1993). To generate P1 DNA for transgenic experiments, a preferred protocol is the protocol described by McCormick et al. (*Genet. Anal. Tech. Appl.*, 11:158-164, 1994). Briefly, *E. coli* (preferably strain NS3529) harboring the P1 plasmid are grown overnight in a suitable broth medium containing 25 µg/ml of kanamycin. The P1 DNA is prepared from the *E. coli* by alkaline lysis using the Qiagen Plasmid Maxi kit (Qiagen, Chatsworth, CA, USA), according to the manufacturer's instructions. The P1 DNA is purified from the bacterial lysate on two Qiagen-tip 500 columns, using the washing and elution buffers contained in the kit. A phenol/chloroform extraction is then performed before precipitating the DNA with 70% ethanol. After solubilizing the DNA in TE (10 mM Tris-HCl, pH 7.4, 1 mM EDTA), the concentration of the DNA is assessed by spectrophotometry.

Yet another viral vector system that is contemplated by the invention consists in the adeno-associated virus (AAV). The adeno-associated virus is a naturally occurring defective virus that requires another virus, such as an adenovirus or a herpes virus, as a helper virus for efficient replication and a productive life cycle (Muzyczka et al., *Current Topics in Microbiol. Immunol.*, 158:97-129, 1992). It is also one of the few viruses that may integrate its DNA into non-dividing cells, and exhibits a high frequency of stable integration (McLaughlin et al., *Am. J. Hum. Genet.*, 59: 561-569, 1989). One advantageous feature of AAV derives from its reduced efficacy for transducing primary cells relative to transformed cells.

The bacterial artificial chromosome (BAC) cloning system (Shizuya et al., *Proc. Natl. Acad. Sci. U.S.A.* 89:8794-8797, 1992) has been developed to stably maintain large fragments of genomic DNA (100-300 kb) in *E. coli*. A preferred BAC vector consists of pBeloBAC11 vector that has been described by Kim et al. (*Genomics*, 34:213-218,1996). BAC libraries are prepared with this vector using size-selected genomic DNA that has been partially digested using enzymes that permit ligation into either the *Bam* HI or *Hind*III sites in the vector. Flanking these cloning sites are T7 and SP6 RNA polymerase transcription initiation sites that can be used to generate end probes by either RNA transcription or PCR methods. After the construction of a BAC library in *E. coli*, BAC DNA is purified from the host cell as a supercoiled circle. Converting these circular molecules into a linear form precedes both size determination and introduction of the BACs into recipient cells. The cloning site is flanked by two *Not* I sites, permitting cloned segments to be excised from the vector by *Not* I digestion. Alternatively, the DNA insert contained in the pBeloBAC11 vector may be linearized by treatment of the BAC vector with the commercially available enzyme lambda terminase that leads to the cleavage at the unique *cos*N site, but this cleavage method results in a full length BAC clone containing both the insert DNA and the BAC sequences.

5. Delivery of the recombinant vectors

In order to effect expression of the polynucleotides and polynucleotide constructs of the invention, these constructs must be delivered into a cell. This delivery may be accomplished *in vitro*, as in laboratory procedures for transforming cell lines, or *in vivo* or *ex vivo*, as in the treatment of certain diseases states. One mechanism is viral infection where the expression construct is encapsidated in an infectious viral particle.

Several non-viral methods for the transfer of polynucleotides into cultured mammalian cells are also contemplated by the present invention, and include, without being limited to, calcium phosphate precipitation (Chen et al., *Proc. Natl. Acad. Sci. USA*, 94:10756-10761, 1987), DEAE-dextran (Gopal, *Mol. Cell. Biol.*, 5:1188-1190, 1985), electroporation (Tur-Kaspa et al., *Mol. Cell. Biol.*, 6:716-718,

1986), direct microinjection (Harland et al., *J. Cell. Biol.* 101:1094-1095, 1985), DNA-loaded liposomes (Nicolau et al., *Biochim. Biophys. Acta.* 721:185-190,1982; Fraley et al., *Natl. Acad. Sci. USA* 76:3348-3352, 1979), and receptor-mediate transfection (Wu and Wu, *J. Biol. Chem.* 262:4429-4432, 1987; Wu and Wu Biochemistry 27:887-892, 1988). Some of these techniques may be successfully adapted for *in vivo* or *ex vivo* use.

Once the expression polynucleotide has been delivered into the cell, it may be stably integrated into the genome of the recipient cell. This integration may be in the cognate location and orientation via homologous recombination (gene replacement) or it may be integrated in a random, non-specific location (gene augmentation). In yet further embodiments, the nucleic acid may be stably maintained in the cell as a separate, episomal segment of DNA. Such nucleic acid segments or "episomes" encode sequences sufficient to permit maintenance and replication independent of or in synchronization with the host cell cycle.

One specific embodiment for a method for delivering a protein or peptide to the interior of a cell of a vertebrate *in vivo* comprises the step of introducing a preparation comprising a physiologically acceptable carrier and a naked polynucleotide operatively coding for the polypeptide of interest into the interstitial space of a tissue comprising the cell, whereby the naked polynucleotide is taken up into the interior of the cell and has a physiological effect. This is particularly applicable for transfer *in vitro* but it may be applied to *in vivo* as well.

Compositions for use *in vitro* and *in vivo* comprising a "naked" polynucleotide are described in PCT application No. WO 90/11092 (Vical Inc.) and in PCT application No. WO 95/11307.

In still another embodiment of the invention, the transfer of a naked polynucleotide of the invention, including a polynucleotide construct of the invention, into cells may be proceeded with a particle bombardment (biolistic), said particles being DNA-coated microprojectiles accelerated to a high velocity allowing them to pierce cell membranes and enter cells without killing them, such as described by Klein et al. (*Nature* 327:70-73, 1987)

In a further embodiment, the polynucleotide of the invention may be entrapped in a liposome (Ghosh and Bacchawat, *Targeting of liposomes to hepatocytes, In: Liver Diseases, Targeted diagnosis and therapy using specific rceptors and ligands,* Marcel Dekeker, New York, 87-104, 1991; Wong et al., *Gene* 10:87-94, 1980; Nicolau et al., *Biochim. Biophys. Acta.* 721:185-190, 1982).

In a specific embodiment, the invention provides a composition for the *in vivo* production of the 12-LO protein or polypeptide described herein. It comprises a naked polynucleotide operatively coding for this polypeptide, in solution in a physiologically acceptable carrier, and suitable for introduction into a tissue to cause cells of the tissue to express the said protein or polypeptide.

The amount of vector to be injected to the desired host organism varies according to the site of injection. As an indicative dose, it will be injected between 0.1 and 100 µg of the vector in an animal body, preferably a mammal body, for example a mouse body.

In another embodiment of the vector according to the invention, it may be introduced in vitro in a host cell, preferably in a host cell previously harvested from the animal to be treated and more preferably a somatic cell such as a muscle cell. In a subsequent step, the cell that has been transformed with the vector coding for the desired 12-LO polypeptide or the desired fragment thereof is reintroduced into the animal body in order to deliver the recombinant protein within the body either locally or systemically.

iii. Host cells

Another object of the invention consists of a host cell that have been transformed or transfected with one of the polynucleotides described therein, and more precisely a polynucleotide either comprising a 12-LO regulatory polynucleotide or the coding sequence of the 12-LO polypeptide having the amino acid sequence of SEQ ID Nos. 653 or 654. Are included host cells that are transformed (prokaryotic cells) or that are transfected (eukaryotic cells) with a recombinant vector such as one of those described above.

Generally, a recombinant host cell of the invention comprises any one of the polynucleotides or the recombinant vectors described therein.

A preferred recombinant host cell according to the invention comprises a polynucleotide selected from the following group of polynucleotides:

a) a purified or isolated nucleic acid encoding a 12-LO polypeptide, or a polypeptide fragment or variant thereof.

b) a purified or isolated nucleic comprising at least 8, preferably at least 15, more preferably at least 25, consecutive nucleotides of the nucleotide sequence SEQ ID No. 651, a nucleotide sequence complementary thereto, or a variant thereof.

c) a purified or isolated nucleic acid comprising at least 8 consecutive nucleotides, preferably at least 15, more preferably at least 25 of the nucleotide sequence SEQ ID No. 652, a nucleotide sequence complementary thereto or a variant thereof.

d) a purified or isolated nucleic acid comprising an exon of the 12-LO gene, a sequence complementary thereto or a fragment or a variant thereof.

e) a purified or isolated nucleic acid comprising a combination of at least two exons of the12-LO gene, or the sequences complementary thereto wherein the polynucleotides are arranged

within the nucleic acid, from the 5' end to the 3'end of said nucleic acid, in the same order than in SEQ ID No. 651.

f) a purified or isolated nucleic acid comprising the nucleotide sequence SEQ ID No. 651 or the sequences complementary thereto or a biologically active fragment thereof.

g) a polynucleotide consisting of:

(1) a nucleic acid comprising a regulatory polynucleotide of SEQ ID No. 651 or the sequences complementary thereto or a biologically active fragment thereof

(2) a polynucleotide encoding a desired polypeptide or nucleic acid.

i) a DNA construct as described previously in the present specification.

Another preferred recombinant cell host according to the present invention is characterized in that its genome or genetic background (including chromosome, plasmids) is modified by the nucleic acid coding for the 12-LO polypeptide of SEQ ID Nos. 653 and 654 or fragments or variants thereof.

Preferred host cells used as recipients for the expression vectors of the invention are the following:

a) Prokaryotic host cells: *Escherichia coli* strains (I.E. DH5-α strain), *Bacillus subtilis*, *Salmonella typhimurium*, and strains from species like *Pseudomonas*, *Streptomyces* and *Staphylococcus*..

b) Eukaryotic host cells: HeLa cells (ATCC N°CCL2; N°CCL2.1; N°CCL2.2), Cv 1 cells (ATCC N°CCL70), COS cells (ATCC N°CRL1650; N°CRL1651), Sf-9 cells (ATCC N°CRL1711), C127 cells (ATCC N° CRL-1804), 3T3 (ATCC N° CRL-6361), CHO (ATCC N° CCL-61), human kidney 293.(ATCC N° 45504; N° CRL-1573) and BHK (ECACC N° 84100501; N° 84111301)

c) Other mammalian host cells:

The 12-LO gene expression in mammalian, and typically human, cells may be rendered defective, or alternatively it may be proceeded with the insertion of a 12-LO genomic or cDNA sequence with the replacement of the 12-LO gene counterpart in the genome of an animal cell by a 12-LO polynucleotide according to the invention. These genetic alterations may be generated by homologous recombination events using specific DNA constructs that have been previously described.

One kind of host cell that may be used is mammalian zygotes, such as murine zygotes. For example, murine zygotes may undergo microinjection with a purified DNA molecule of interest, such as a purified DNA molecule that has previously been adjusted to a concentration range from 1 ng/ml (for BAC inserts) 3 ng/μl (for P1 bacteriophage inserts) in 10 mM Tris-HCl, pH 7.4, 250 μM EDTA containing 100 mM NaCl, 30 μM spermine, and70 μM spermidine. When the DNA to be microinjected

is relatively large, polyamines and high salt concentrations can be used to avoid mechanical breakage of this DNA, as described by Schedl et al (*Nucleic Acids Res.* 21:4783-4787, 1993).

Anyone of the polynucleotides of the invention, including the DNA constructs described herein, may be introduced in an embryonic stem (ES) cell line, preferably a mouse ES cell line. ES cell lines are derived from pluripotent, uncommitted cells of the inner cell mass of pre-implantation blastocysts. Preferred ES cell lines are the following: ES-E14TG2a (ATCC n° CRL-1821), ES-D3 (ATCC n° CRL1934 and n° CRL-11632), YS001 (ATCC n° CRL-11776), 36.5 (ATCC n° CRL-11116). To maintain ES cells in an uncommitted state, they are cultured in the presence of growth inhibited feeder cells, which provide the appropriate signals to preserve this embryonic phenotype and serve as a matrix for ES cell adherence. Preferred feeder cells consist of primary embryonic fibroblasts that are established from tissue of day 13- day 14 embryos of virtually any mouse strain, that are maintained in culture, such as described by Abbondanzo et al. (*Methods in Enzymology*, Academic Press, NewYork, 803-823, 1993), and are inhibited in growth by irradiation, such as described by Robertson ("Embryo-Derived StemCell Lines," *E.J. Robertson Ed..Teratocarcinomas and Embrionic Stem Cells: A Practical Approach.* IRL Press, Oxford, 71, 1987), or by the presence of an inhibitory concentration of LIF, such as described by Pease and Williams (*Exp. Cell. Res.* 190:09-211, 1990).

The constructs in the host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence.

Following transformation of a suitable host and growth of the host to an appropriate cell density, the selected promoter is induced by appropriate means, such as temperature shift or chemical induction, and cells are cultivated for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in the expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Such methods are well known by the skill artisan.

iv. Transgenic animals

The terms "transgenic animals" or "host animals" used herein designate animals that have their genome genetically and artificially manipulated so as to include one of the nucleic acids according to the invention. Preferred animals are non-human mammals and include those belonging to a genus selected from *Mus* (e.g. mice), *Rattus* (e.g. rats) and *Oryctogalus* (e.g. rabbits) which have their genome artificially and genetically altered by the insertion of a nucleic acid according to the invention.

The transgenic animals of the invention all include within a plurality of their cells a cloned recombinant or synthetic DNA sequence, more specifically one of the purified or isolated nucleic acids comprising a 12-LO coding sequence, a 12-LO regulatory polynucleotide or a DNA sequence encoding an antisense polynucleotide such as described in the present specification.

Preferred transgenic animals according to the invention contain in their somatic cells and/or in their germ line cells a polynucleotide selected from the following group of polynucleotides:

a) a purified or isolated nucleic acid encoding a 12-LO polypeptide, or a polypeptide fragment or variant thereof.

b) a purified or isolated nucleic comprising at least 8, preferably at least 15, more preferably at least 25, consecutive nucleotides of the nucleotide sequence SEQ ID No. 651, a nucleotide sequence complementary thereto.

c) a purified or isolated nucleic acid comprising at least 8 consecutive nucleotides, preferably at least 15, more preferably at least 25 of the nucleotide sequence SEQ ID No. 652, a nucleotide sequence complementary thereto.

d) a purified or isolated nucleic acid comprising an exon of the 12-LO gene, a sequence complementary thereto or a fragment or a variant thereof.

e) a purified or isolated nucleic acid comprising a combination of at least two exons of the 12-LO gene, or the sequences complementary thereto wherein the polynucleotides are arranged within the nucleic acid, from the 5' end to the 3'end of said nucleic acid, in the same order than in SEQ ID No. 651.

f) a purified or isolated nucleic acid comprising the nucleotide sequence SEQ ID No. 651 or the sequences complementary thereto or a biologically active fragment thereof.

g) a polynucleotide consisting of:

(1) a nucleic acid comprising a regulatory polynucleotide of SEQ ID No. 651 or the sequences complementary thereto or a biologically active fragment thereof

2) a polynucleotide encoding a desired polypeptide or nucleic acid.

i) a DNA construct as described previously in the present specification.

The transgenic animals of the invention thus contain specific sequences of exogenous genetic material such as the nucleotide sequences described above in detail.

In a first preferred embodiment, these transgenic animals may be good experimental models in order to study the diverse pathologies related to cell differentiation, in particular concerning the transgenic animals within the genome of which has been inserted one or several copies of a polynucleotide encoding a native 12-LO protein, or alternatively a mutant 12-LO protein.

In a second preferred embodiment, these transgenic animals may express a desired polypeptide of interest under the control of the regulatory polynucleotides of the 12-LO gene, leading to good yields in the synthesis of this protein of interest, and eventually a tissue specific expression of this protein of interest.

The design of the transgenic animals of the invention may be made according to the conventional techniques well known for one skilled in the art. For more details regarding the production of transgenic animals, and specifically transgenic mice, one may refer to US Patents Nos. 4,873,191, issued Oct.10, 1989, 5,464,764 issued Nov. 7, 1995 and 5,789,215, issued Aug. 4, 1998.

Transgenic animals of the present invention are produced by the application of procedures which result in an animal with a genome that has incorporated exogenous genetic material. The procedure involves obtaining the genetic material, or a portion thereof, which encodes either a 12-LO coding sequence, a 12-LO regulatory polynucleotide or a DNA sequence encoding a 12-LO antisense polynucleotide such as described in the present specification.

A recombinant polynucleotide of the invention is inserted into an embryonic or ES stem cell line. The insertion is preferably made using electroporation, such as described by Thomas et al. (*Cell* 51:503-512, 1987). The cells subjected to electroporation are screened (e.g. by selection via selectable markers, by PCR or by Southern blot analysis) to find positive cells which have integrated the exogenous recombinant polynucleotide into their genome, preferably via an homologous recombination event. An illustrative positive-negative selection procedure that may be used according to the invention is described by Mansour et al. (*Nature* 336:348-352, 1988).

Then, the positive cells are isolated, cloned and injected into 3.5 days old blastocysts from mice, such as described by Bradley ("Production and Analysis of Chimaeric Mice," *E.J. Robertson (Ed.), Teratocarcinomas and embryonic stem cells: A practical approach* IRL Press, Oxford, 113, 1987). The blastocysts are then inserted into a female host animal and allowed to grow to term.

Alternatively, the positive ES cells are brought into contact with embryos at the 2.5 days old 8-16 cell stage (morulae) such as described by Wood et al. (*Proc. Natl. Acad. Sci. U.S.A.* 90:4582-4585, 1993) or by Nagy et al. (*Proc. Natl. Acad. Sci. USA*. 90: 8424-8428, 1993), the ES cells being internalized to colonize extensively the blastocyst including the cells which will give rise to the germ line.

The offspring of the female host are tested to determine which animals are transgenic e.g. include the inserted exogenous DNA sequence and which are wild-type. Thus, the present invention also concerns a transgenic animal containing a nucleic acid, a recombinant expression vector or a recombinant host cell according to the invention.

A further object of the invention consists of recombinant host cells obtained from a transgenic animal described herein.

Recombinant cell lines may be established *in vitro* from cells obtained from any tissue of a transgenic animal according to the invention, for example by transfection of primary cell cultures with vectors expressing *onc*-genes such as SV40 large T antigen, as described by Chou (*Mol. Endocrinol.* 3:1511-1514, 1989) and Shay et al. (*Biochem. Biophys. Acta.* 1072:1-7, 1991).

I.E. 12-Lipoxygenase Polypeptides

The term "12-LO polypeptides" is used herein to embrace all of the proteins and polypeptides of the present invention. Also forming part of the invention are polypeptides encoded by the polynucleotides of the invention, as well as fusion polypeptides comprising such polypeptides. The invention embodies 12-LO proteins from humans, including isolated or purified 12-LO proteins consisting, consisting essentially, or comprising the sequence of SEQ ID Nos. 653 and 654.

Biallelic markers are associated with amino acid substitutions in the polypeptide sequence of 12-LO. It should be noted the 12-LO proteins of the invention are based on the naturally-occurring variants of the amino acid sequence of human 12-LO; wherein the Arg residue of amino acid position 189 has been replaced with a His residue (biallelic marker 10-346-141), the Asp residue of amino acid position 225 has been replaced with a His residue (biallelic marker 10-347-111), the Arg residue of amino acid position 243 has been replaced with a Cys residue (biallelic marker 10-347-165), the Gln residue of amino acid position 261 has been replaced with an Arg residue (biallelic marker 10-347-220), the Ser residue of amino acid position 322 has been replaced with a Asn residue (biallelic marker 10-349-97), the Pro residue of amino acid position 337 has been replaced with an Arg residue (biallelic marker 10-349-142), the Thr residue of amino acid position 568 has been replaced with an Asn residue (biallelic marker 10-340-112) and wherein the Met residue of amino acid position 574 has been replaced with a Lys residue (biallelic marker 10-340-112). Variant proteins and the fragments thereof which contain amino acid position 189 are collectively referred to herein as "189-His variants." Variant proteins and the fragments thereof which contain amino acid position 225 are collectively referred to herein as "225-His variants." Variant proteins and the fragments thereof which, contain amino acid position 243, are collectively referred to herein as "243-Cys variants." Variant proteins and the fragments thereof which contain amino acid position 261 are collectively referred to herein as "261-Arg variants." Variant proteins and the fragments thereof which contain amino acid position 322 are collectively referred to herein as "322-Asn variants." Variant proteins and the fragments thereof which contain amino acid position 337 are collectively referred to herein as "337-Arg variants." Variant proteins and the fragments thereof which contain amino acid position 568 are collectively referred to

herein as "568-Asn variants." Variant proteins and the fragments thereof which contain amino acid position 574 are collectively referred to herein as "574-Lys variants." In each of these amino acid substitutions the original residue is replaced by a non-equivalent amino acid presenting different chemical properties. Therefore, these substitutions cause alterations in the activity, specificity and function of the 12-LO enzyme.

One allele of biallelic marker 10-349-216 is associated with the deletion of a Leu residue at amino acid position 362 of SEQ ID No. 653. 12-LO polypeptides of the present invention also include 12-LO polypeptides wherein the Leu residue at amino acid position 362 of SEQ ID No. 653 has been deleted.

One allele of biallelic marker 10-343-231 is associated with a frameshift in the open reading frame of the 12-LO gene leading to the expression of the variant 12-LO polypepetide of SEQ ID No. 654.

The present invention embodies isolated, purified, and recombinant polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No. 653, wherein said contiguous span comprises at least one amino acid position selected from the group consisting of: an His residue et amino acid position 189, an His residue at amino acid position 225, a Cys residue at amino acid position 243, an Arg residue at amino acid position 261, an Asn residue at amino acid position 322, an Arg residue at amino acid position 337, a Asn residue at amino acid position 362, an Asn at amino acid position 568 and a Lys residue at amino acid position 574.

The present invention further provides isolated, purified, and recombinant polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No. 654, wherein said contiguous span comprises at least one of amino acid positions 110-131 of SEQ ID No. 654.

The present invention further embodies isolated, purified, and recombinant polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No. 653, wherein said contiguous span comprises a Leu residue at amino acid position 389 of SEQ ID No. 653.

In other preferred embodiments the contiguous stretch of amino acids comprises the site of a mutation or functional mutation, including a deletion, addition, swap or truncation of the amino acids in the 12-LO protein sequence.

12-LO proteins are preferably isolated from human or mammalian tissue samples or expressed from human or mammalian genes. The 12-LO polypeptides of the invention can be made using routine expression methods known in the art. The polynucleotide encoding the desired polypeptide is ligated

into an expression vector suitable for any convenient host. Both eukaryotic and prokaryotic host systems are used in forming recombinant polypeptides. The polypeptide is then isolated from lysed cells or from the culture medium and purified to the extent needed for its intended use. Purification is by any technique known in the art, for example, differential extraction, salt fractionation, chromatography, centrifugation, and the like. See, for example, *Methods in Enzymology* for a variety of methods for purifying proteins.

In addition, shorter protein fragments are produced by chemical synthesis. Alternatively the proteins of the invention are extracted from cells or tissues of humans or non-human animals. Methods for purifying proteins are known in the art, and include the use of detergents or chaotropic agents to disrupt particles followed by differential extraction and separation of the polypeptides by ion exchange chromatography, affinity chromatography, sedimentation according to density, and gel electrophoresis.

Any 12-LO cDNA, including SEQ ID No. 652, is used to express 12-LO proteins and polypeptides. The nucleic acid encoding the 12-LO protein or polypeptide to be expressed is operably linked to a promoter in an expression vector using conventional cloning technology. The 12-LO insert in the expression vector may comprise the full coding sequence for the 12-LO protein or a portion thereof.

The expression vector is any of the mammalian, yeast, insect or bacterial expression systems known in the art. Commercially available vectors and expression systems are available from a variety of suppliers including Genetics Institute (Cambridge, MA), Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence is optimized for the particular expression organism in which the expression vector is introduced, as explained by Hatfield, et al., U.S. Patent No. 5,082,767.

In one embodiment, the entire coding sequence of the 12-LO cDNA through the poly A signal of the cDNA is operably linked to a promoter in the expression vector. Alternatively, if the nucleic acid encoding a portion of the 12-LO protein lacks a methionine to serve as the initiation site, an initiating methionine can be introduced next to the first codon of the nucleic acid using conventional techniques. Similarly, if the insert from the 12-LO cDNA lacks a poly A signal, this sequence can be added to the construct by, for example, splicing out the Poly A signal from pSG5 (Stratagene) using BglI and SalI restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene). pXT1 contains the LTRs and a portion of the gag gene from Moloney Murine Leukemia Virus. The position of the LTRs in the construct allow efficient stable transfection. The vector includes the Herpes Simplex Thymidine Kinase promoter and the selectable neomycin gene. The nucleic acid encoding the 12-LO protein or a portion thereof is obtained by PCR from a bacterial vector containing the 12-LO cDNA of SEQ ID No. 652 using oligonucleotide primers complementary to the 12-LO cDNA or

portion thereof and containing restriction endonuclease sequences for Pst I incorporated into the 5'primer and BglII at the 5' end of the corresponding cDNA 3' primer, taking care to ensure that the sequence encoding the 12-LO protein or a portion thereof is positioned properly with respect to the poly A signal. The purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with Bgl II, purified and ligated to pXT1, now containing a poly A signal and digested with BglII.

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. Positive transfectants are selected after growing the transfected cells in 600ug/ml G418 (Sigma, St. Louis, Missouri).

Alternatively, the nucleic acids encoding the 12-LO protein or a portion thereof is cloned into pED6dpc2 (Genetics Institute, Cambridge, MA). The resulting pED6dpc2 constructs is transfected into a suitable host cell, such as COS 1 cells. Methotrexate resistant cells are selected and expanded.

The above procedures may also be used to express a mutant 12-LO protein responsible for a detectable phenotype or a portion thereof.

The expressed proteins are purified using conventional purification techniques such as ammonium sulfate precipitation or chromatographic separation based on size or charge. The protein encoded by the nucleic acid insert may also be purified using standard immunochromatography techniques. In such procedures, a solution containing the expressed 12-LO protein or portion thereof, such as a cell extract, is applied to a column having antibodies against the 12-LO protein or portion thereof is attached to the chromatography matrix. The expressed protein is allowed to bind the immunochromatography column. Thereafter, the column is washed to remove non-specifically bound proteins. The specifically bound expressed protein is then released from the column and recovered using standard techniques.

To confirm expression of the 12-LO protein or a portion thereof, the proteins expressed from host cells containing an expression vector containing an insert encoding the 12-LO protein or a portion thereof can be compared to the proteins expressed in host cells containing the expression vector without an insert. The presence of a band in samples from cells containing the expression vector with an insert which is absent in samples from cells containing the expression vector without an insert indicates that the 12-LO protein or a portion thereof is being expressed. Generally, the band will have the mobility expected for the 12-LO protein or portion thereof. However, the band may have a mobility different than that expected as a result of modifications such as glycosylation, ubiquitination, or enzymatic cleavage.

Antibodies capable of specifically recognizing the expressed 12-LO protein or a portion thereof, are described below.

If antibody production is not possible, the nucleic acids encoding the 12-LO protein or a portion thereof is incorporated into expression vectors designed for use in purification schemes employing chimeric polypeptides. In such strategies the nucleic acid encoding the 12-LO protein or a portion thereof is inserted in frame with the gene encoding the other half of the chimera. The other half of the chimera is β-globin or a nickel binding polypeptide encoding sequence. A chromatography matrix having antibody to β-globin or nickel attached thereto is then used to purify the chimeric protein. Protease cleavage sites is engineered between the β-globin gene or the nickel binding polypeptide and the 12-LO protein or portion thereof. Thus, the two polypeptides of the chimera are separated from one another by protease digestion.

One useful expression vector for generating β-globin chimerics is pSG5 (Stratagene), which encodes rabbit β-globin. Intron II of the rabbit β-globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis et al., (*Basic Methods in Molecular Biology*, L.G. Davis, M.D. Dibner, and J.F. Battey, ed., Elsevier Press, NY, 1986) and many of the methods are available from Stratagene, Life Technologies, Inc., or Promega. Polypeptide may additionally be produced from the construct using in vitro translation systems such as the In vitro Express™ Translation Kit (Stratagene).

I.F. Production Of Antibodies Against 12-lipoxygenase Polypeptides

Any 12-LO polypeptide or whole protein may be used to generate antibodies capable of specifically binding to expressed 12-LO protein or fragments thereof as described. The antibody compositions of the invention are capable of specifically binding to the 189-His variant of the 12-LO protein or, to the 225-His variant of the 12-LO protein or, to the 243-Cys variant of the 12-LO protein or, to the 261-Arg variant of the 12-LO protein or, to the 322-Asn variant of the 12-LO or, to the 337-Arg variant of the 12-LO protein or to the 574-Lys variant of the 12-LO protein. A preferred embodiment of the invention encompasses isolated or purified antibody compositions capable of selectively binding, or which are capable of binding to an epitope-containing fragment of a polypeptide of the invention, wherein said epitope comprises at least one amino acid position selected from the group consisting of an His residue et amino acid position 189, an His residue at amino acid position 225, a Cys residue at amino acid position 243, an Arg residue at amino acid position 261, an Asn residue at amino acid position 322, an Arg residue at amino acid position 337, a Asn residue at amino acid position 362, an Asn at amino acid position 568 and a Lys residue at amino acid position 574. For an antibody composition to specifically bind to these 12-LO variants it must demonstrate at least a 5%, 10%, 15%, 20%, 25%, 50%, or 100% greater binding affinity for full length 189-His, 225-His, 243-Cys, 261-Arg, 322- Asn, 337-Arg or 574-Lys variants in an ELISA, RIA, or other antibody-based binding

assay than to full length 12-LO proteins which have the alternative amino acid specified in SEQ ID No. 653. Affinity of the antibody composition for the epitope can further be determined by preparing competitive binding curves, as described, for example, by Fisher, D., (Manual of Clinical Immunology, 2nd Ed. (Rose and Friedman,Eds.) *Amer. Soc. For Microbiol.*, Washington, D.C., Ch. 42, 1980).

Other preferred antibody compositions of the invention are capable of specifically binding to amino acid positions 110-131 of SEQ ID No. 654.

The present invention also contemplates the use of polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 50, or 100 amino acids of a 12-LO polypeptide of SEQ ID No. 653 in the manufacture of antibodies, wherein said contiguous span comprises at least one amino acid position selected from the group consisting of: an His residue et amino acid position 189, an His residue at amino acid position 225, a Cys residue at amino acid position 243, an Arg residue at amino acid position 261, an Asn residue at amino acid position 322, an Arg residue at amino acid position 337, a Asn residue at amino acid position 362, an Asn at amino acid position 568 and a Lys residue at amino acid position 574.

In a preferred embodiment such polypeptides are useful in the manufacture of antibodies to detect the presence and absence of the 189-His, 225-His, 243-Cys, 261-Arg, 322- Asn, 337-Arg, 568-Asn, or 574-Lys variant.

The present invention further encompasses the use of isolated, purified, and recombinant polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No. 654, wherein said contiguous span comprises at least one of amino acid positions 110-131 of SEQ ID No. 654.

In a preferred embodiment such polypeptides are useful in the manufacture of antibodies to detect the presence and absence of amino acid positions 110-131 of SEQ ID No. 654.

Non-human animals or mammals, whether wild-type or transgenic, which express a different species of 12-LO than the one to which antibody binding is desired, and animals which do not express 12-LO (i.e. an 12-LO knock out animal as described in herein) are particularly useful for preparing antibodies. 12-LO knock out animals will recognize all or most of the exposed regions of 12-LO as foreign antigens, and therefore produce antibodies with a wider array of 12-LO epitopes. Moreover, smaller polypeptides with only 10 to 30 amino acids may be useful in obtaining specific binding to the 189-His, 225-His, 243-Cys, 261-Arg, 322-Asn, 337-Arg, 568-Asn, or 574-Lys variants. In addition, the humoral immune system of animals which produce a species of 12-LO that resembles the antigenic sequence will preferentially recognize the differences between the animal's native 12-LO species and the antigen sequence, and produce antibodies to these unique sites in the antigen sequence. Such a

technique will be particularly useful in obtaining antibodies that specifically bind to the 189-His, 225-His, 243-Cys, 261-Arg, 322- Asn, 337-Arg, 568-Asn, or 574-Lys variants. The preparation of antibody compositions is further described in Example 6.

Antibody preparations prepared according to the present invention are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. The antibodies may also be used in therapeutic compositions for killing cells expressing the protein or reducing the levels of the protein in the body. The antibodies of the invention may be labeled, either by a radioactive, a fluorescent or an enzymatic label. Consequently, the invention is also directed to a method for detecting specifically the presence of a variant 12-LO polypeptide according to the invention in a biological sample, said method comprising the following steps : a) bringing into contact the biological sample with a polyclonal or monoclonal antibody that specifically binds a variant 12-LO polypeptide or to a peptide fragment or variant thereof; and b) detecting the antigen-antibody complex formed. The invention also concerns a diagnostic kit for detecting *in vitro* the presence of a variant 12-LO polypeptide according to the present invention in a biological sample, wherein said kit comprises:

a) a polyclonal or monoclonal antibody that specifically binds a variant 12-LO polypeptide or to a peptide fragment or variant thereof, optionally labeled;

b) a reagent allowing the detection of the antigen-antibody complexes formed, said reagent carrying optionally a label, or being able to be recognized itself by a labeled reagent, more particularly in the case when the above-mentioned monoclonal or polyclonal antibody is not labeled by itself.

II. Methods for *De Novo* Identification of Biallelic Markers

Large fragments of human DNA, carrying genes of interest involved in arachidonic acid metabolism; were cloned, sequenced and screened for biallelic markers. Biallelic markers within the candidate genes themselves as well as markers located on the same genomic fragment were identified. It will be clear to one of skill in the art that large fragments of human genomic DNA may be obtained from any appropriate source and may be cloned into a number of suitable vectors.

In a preferred embodiment of the invention, BAC (Bacterial Artificial Chromosomes) vectors were used to construct DNA libraries covering the entire human genome. Specific amplification primers were designed for each candidate gene and the BAC library was screened by PCR until there was at least one positive BAC clone per candidate gene. Genomic sequence, screened for biallelic markers, was generated by sequencing ends of BAC subclones. Details of a preferred embodiment are provided in Example 1. As a preferred alternative to sequencing the ends of an adequate number of

BAC subclones, high throughput deletion-based sequencing vectors, which allow the generation of a high quality sequence information covering fragments of about 6kb, may be used. Having sequence fragments longer than 2.5 or 3kb enhances the chances of identifying biallelic markers therein. Methods of constructing and sequencing a nested set of deletions are disclosed in the related U.S. Patent Application entitled "High Throughput DNA Sequencing Vector" (Serial No. 09/058,746).

In another embodiment of the invention, genomic sequences of candidate genes were available in public databases allowing direct screening for biallelic markers.

Any of a variety of methods can be used to screen a genomic fragment for single nucleotide polymorphisms such as differential hybridization with oligonucleotide probes, detection of changes in the mobility measured by gel electrophoresis or direct sequencing of the amplified nucleic acid. A preferred method for identifying biallelic markers involves comparative sequencing of genomic DNA fragments from an appropriate number of unrelated individuals.

In a first embodiment, DNA samples from unrelated individuals are pooled together, following which the genomic DNA of interest is amplified and sequenced. The nucleotide sequences thus obtained are then analyzed to identify significant polymorphisms. One of the major advantages of this method resides in the fact that the pooling of the DNA samples substantially reduces the number of DNA amplification reactions and sequencing reactions, which must be carried out. Moreover, this method is sufficiently sensitive so that a biallelic marker obtained thereby usually demonstrates a sufficient frequency of its less common allele to be useful in conducting association studies. Usually, the frequency of the least common allele of a biallelic marker identified by this method is at least 10%.

In a second embodiment, the DNA samples are not pooled and are therefore amplified and sequenced individually. This method is usually preferred when biallelic markers need to be identified in order to perform association studies within candidate genes. Preferably, highly relevant gene regions such as promoter regions or exon regions may be screened for biallelic markers. A biallelic marker obtained using this method may show a lower degree of informativeness for conducting association studies, e.g. if the frequency of its less frequent allele may be less than about 10%. Such a biallelic marker will however be sufficiently informative to conduct association studies and it will further be appreciated that including less informative biallelic markers in the genetic analysis studies of the present invention, may allow in some cases the direct identification of causal mutations, which may, depending on their penetrance, be rare mutations.

The following is a description of the various parameters of a preferred method used by the inventors for the identification of the biallelic markers of the present invention.

II.A. Genomic DNA samples

The genomic DNA samples from which the biallelic markers of the present invention are generated are preferably obtained from unrelated individuals corresponding to a heterogeneous population of known ethnic background. The number of individuals from whom DNA samples are obtained can vary substantially, preferably from about 10 to about 1000, more preferably from about 50 to about 200 individuals. Usually, DNA samples are collected from at least about 100 individuals in order to have sufficient polymorphic diversity in a given population to identify as many markers as possible and to generate statistically significant results.

As for the source of the genomic DNA to be subjected to analysis, any test sample can be foreseen without any particular limitation. These test samples include biological samples, which can be tested by the methods of the present invention described herein, and include human and animal body fluids such as whole blood, serum, plasma, cerebrospinal fluid, urine, lymph fluids, and various external secretions of the respiratory, intestinal and genitourinary tracts, tears, saliva, milk, white blood cells, myelomas and the like; biological fluids such as cell culture supernatants; fixed tissue specimens including tumor and non-tumor tissue and lymph node tissues; bone marrow aspirates and fixed cell specimens. The preferred source of genomic DNA used in the present invention is from peripheral venous blood of each donor. Techniques to prepare genomic DNA from biological samples are well known to the skilled technician. Details of a preferred embodiment are provided in Example 1. A person skilled in the art can choose to amplify pooled or unpooled DNA samples.

II.B. DNA Amplification

The identification of biallelic markers in a sample of genomic DNA may be facilitated through the use of DNA amplification methods. DNA samples can be pooled or unpooled for the amplification step. DNA amplification techniques are well known to those skilled in the art. Various methods to amplify DNA fragments carrying biallelic markers are further described hereinafter in III.B. The PCR technology is the preferred amplification technique used to identify new biallelic markers.

In a first embodiment, biallelic markers are identified using genomic sequence information generated by the inventors. Genomic DNA fragments, such as the inserts of the BAC clones described above, are sequenced and used to design primers for the amplification of 500 bp fragments. These 500 bp fragments are amplified from genomic DNA and are scanned for biallelic markers. Primers may be designed using the OSP software (Hillier L. and Green P., *Methods Appl.* 1: 124-8, 1991). All primers may contain, upstream of the specific target bases, a common oligonucleotide tail that serves as a sequencing primer. Those skilled in the art are familiar with primer extensions, which can be used for these purposes.

In another embodiment of the invention, genomic sequences of candidate genes are available in public databases allowing direct screening for biallelic markers. Preferred primers, useful for the amplification of genomic sequences encoding the candidate genes, focus on promoters, exons and splice sites of the genes. A biallelic marker present in these functional regions of the gene has a higher probability to be a causal mutation.

Preferred primers include those disclosed in Figure 8.

II.C. Sequencing Of Amplified Genomic DNA And Identification Of Single Nucleotide Polymorphisms

The amplification products generated as described above, are then sequenced using any method known and available to the skilled technician. Methods for sequencing DNA using either the dideoxy-mediated method (Sanger method) or the Maxam-Gilbert method are widely known to those of ordinary skill in the art. Such methods are for example disclosed in Maniatis et al. (Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, 2nd Edition, 1989). Alternative approaches include hybridization to high-density DNA probe arrays as described in Chee et al. (*Science* 274:610, 1996).

Preferably, the amplified DNA is subjected to automated dideoxy terminator sequencing reactions using a dye-primer cycle sequencing protocol. The products of the sequencing reactions are run on sequencing gels and the sequences are determined using gel image analysis. The polymorphism search is based on the presence of superimposed peaks in the electrophoresis pattern resulting from different bases occurring at the same position. Because each dideoxy terminator is labeled with a different fluorescent molecule, the two peaks corresponding to a biallelic site present distinct colors corresponding to two different nucleotides at the same position on the sequence. However, the presence of two peaks can be an artifact due to background noise. To exclude such an artifact, the two DNA strands are sequenced and a comparison between the peaks is carried out. In order to be registered as a polymorphic sequence, the polymorphism has to be detected on both strands.

The above procedure permits those amplification products, which contain biallelic markers to be identified. The detection limit for the frequency of biallelic polymorphisms detected by sequencing pools of 100 individuals is approximately 0.1 for the minor allele, as verified by sequencing pools of known allelic frequencies. However, more than 90% of the biallelic polymorphisms detected by the pooling method have a frequency for the minor allele higher than 0.25. Therefore, the biallelic markers selected by this method have a frequency of at least 0.1 for the minor allele and less than 0.9 for the major allele. Preferably at least 0.2 for the minor allele and less than 0.8 for the major allele, more preferably at least 0.3 for the minor allele and less than 0.7 for the major allele, thus a heterozygosity rate higher than 0.18, preferably higher than 0.32, more preferably higher than 0.42.

In another embodiment, biallelic markers are detected by sequencing individual DNA samples, the frequency of the minor allele of such a biallelic marker may be less than 0.1.

The markers carried by the same fragment of genomic DNA, such as the insert in a BAC clone, need not necessarily be ordered with respect to one another within the genomic fragment to conduct association studies. However, in some embodiments of the present invention, the order of biallelic markers carried by the same fragment of genomic DNA are determined.

II.D. Validation of the Biallelic Markers of the Present Invention

The polymorphisms are evaluated for their usefulness as genetic markers by validating that both alleles are present in a population. Validation of the biallelic markers is accomplished by genotyping a group of individuals by a method of the invention and demonstrating that both alleles are present. Microsequencing is a preferred method of genotyping alleles. The validation by genotyping step may be performed on individual samples derived from each individual in the group or by genotyping a pooled sample derived from more than one individual. The group can be as small as one individual if that individual is heterozygous for the allele in question. Preferably the group contains at least three individuals, more preferably the group contains five or six individuals, so that a single validation test will be more likely to result in the validation of more of the biallelic markers that are being tested. It should be noted, however, that when the validation test is performed on a small group it may result in a false negative result if as a result of sampling error none of the individuals tested carries one of the two alleles. Thus, the validation process is less useful in demonstrating that a particular initial result is an artifact, than it is at demonstrating that there is a *bona fide* biallelic marker at a particular position in a sequence. For an indication of whether a particular biallelic marker has been validated see Figure 2. All of the genotyping, haplotyping, association, and interaction study methods of the invention may optionally be performed solely with validated biallelic markers.

II.E. Evaluation of the Frequency of the Biallelic Markers of the Present Invention

The validated biallelic markers are further evaluated for their usefulness as genetic markers by determining the frequency of the least common allele at the biallelic marker site. The determination of the least common allele is accomplished by genotyping a group of individuals by a method of the invention and demonstrating that both alleles are present. This determination of frequency by genotyping step may be performed on individual samples derived from each individual in the group or by genotyping a pooled sample derived from more than one individual. The group must be large enough to be representative of the population as a whole. Preferably the group contains at least 20 individuals, more preferably the group contains at least 50 individuals, most preferably the group

contains at least 100 individuals. Of course the larger the group the greater the accuracy of the frequency determination because of reduced sampling error. For an indication of the frequency for the less common allele of a particular biallelic marker of the invention see Figure 2. A biallelic marker wherein the frequency of the less common allele is 30% or more is termed a "high quality biallelic marker." All of the genotyping, haplotyping, association, and interaction study methods of the invention may optionally be performed solely with high quality biallelic markers.

III. Methods Of Genotyping an Individual for Biallelic Markers

Methods are provided to genotype a biological sample for one or more biallelic markers of the present invention, all of which may be performed *in vitro*. Such methods of genotyping comprise determining the identity of a nucleotide at an eicosanoid-related biallelic marker by any method known in the art. These methods find use in genotyping case-control populations in association studies as well as individuals in the context of detection of alleles of biallelic markers which, are known to be associated with a given trait, in which case both copies of the biallelic marker present in individual's genome are determined so that an individual may be classified as homozygous or heterozygous for a particular allele.

These genotyping methods can be performed nucleic acid samples derived from a single individual or pooled DNA samples.

Genotyping can be performed using similar methods as those described above for the identification of the biallelic markers, or using other genotyping methods such as those further described below. In preferred embodiments, the comparison of sequences of amplified genomic fragments from different individuals is used to identify new biallelic markers whereas microsequencing is used for genotyping known biallelic markers in diagnostic and association study applications.

III.A. Source of DNA for Genotyping

Any source of nucleic acids, in purified or non-purified form, can be utilized as the starting nucleic acid, provided it contains or is suspected of containing the specific nucleic acid sequence desired. DNA or RNA may be extracted from cells, tissues, body fluids and the like as described above in II.A. While nucleic acids for use in the genotyping methods of the invention can be derived from any mammalian source, the test subjects and individuals from which nucleic acid samples are taken are generally understood to be human.

III.B. Amplification Of DNA Fragments Comprising Biallelic Markers

Methods and polynucleotides are provided to amplify a segment of nucleotides comprising one or more biallelic marker of the present invention. It will be appreciated that amplification of DNA fragments comprising biallelic markers may be used in various methods and for various purposes and is not restricted to genotyping. Nevertheless, many genotyping methods, although not all, require the previous amplification of the DNA region carrying the biallelic marker of interest. Such methods specifically increase the concentration or total number of sequences that span the biallelic marker or include that site and sequences located either distal or proximal to it. Diagnostic assays may also rely on amplification of DNA segments carrying a biallelic marker of the present invention.

Amplification of DNA may be achieved by any method known in the art. The established PCR (polymerase chain reaction) method or by developments thereof or alternatives. Amplification methods which can be utilized herein include but are not limited to Ligase Chain Reaction (LCR) as described in EP A 320 308 and EP A 439 182, Gap LCR (Wolcott, M.J., Clin. Mcrobiol. Rev. 5:370-386), the so-called "NASBA" or "3SR" technique described in Guatelli J.C. et al. (*Proc. Natl. Acad. Sci. USA* 87:1874-1878, 1990) and in Compton J. (*Nature* 350:91-92, 1991), Q-beta amplification as described in European Patent Application no 4544610, strand displacement amplification as described in Walker et al. (*Clin. Chem.* 42:9-13, 1996) and EP A 684 315 and, target mediated amplification as described in PCT Publication WO 9322461.

LCR and Gap LCR are exponential amplification techniques, both depend on DNA ligase to join adjacent primers annealed to a DNA molecule. In Ligase Chain Reaction (LCR), probe pairs are used which include two primary (first and second) and two secondary (third and fourth) probes, all of which are employed in molar excess to target. The first probe hybridizes to a first segment of the target strand and the second probe hybridizes to a second segment of the target strand, the first and second segments being contiguous so that the primary probes abut one another in 5' phosphate-3'hydroxyl relationship, and so that a ligase can covalently fuse or ligate the two probes into a fused product. In addition, a third (secondary) probe can hybridize to a portion of the first probe and a fourth (secondary) probe can hybridize to a portion of the second probe in a similar abutting fashion. Of course, if the target is initially double stranded, the secondary probes also will hybridize to the target complement in the first instance. Once the ligated strand of primary probes is separated from the target strand, it will hybridize with the third and fourth probes which can be ligated to form a complementary, secondary ligated product. It is important to realize that the ligated products are functionally equivalent to either the target or its complement. By repeated cycles of hybridization and ligation, amplification of the target sequence is achieved. A method for multiplex LCR has also been described (WO 9320227). Gap LCR (GLCR) is a version of LCR where the probes are not adjacent but are separated by 2 to 3 bases.

For amplification of mRNAs, it is within the scope of the present invention to reverse transcribe mRNA into cDNA followed by polymerase chain reaction (RT-PCR); or, to use a single enzyme for both steps as described in U.S. Patent No. 5,322,770 or, to use Asymmetric Gap LCR (RT-AGLCR) as described by Marshall R.L. et al. (*PCR Methods and Applications* 4:80-84, 1994). AGLCR is a modification of GLCR that allows the amplification of RNA.

Some of these amplification methods are particularly suited for the detection of single nucleotide polymorphisms and allow the simultaneous amplification of a target sequence and the identification of the polymorphic nucleotide as it is further described in IIIC.

The PCR technology is the preferred amplification technique used in the present invention. A variety of PCR techniques are familiar to those skilled in the art. For a review of PCR technology, see Molecular Cloning to Genetic Engineering White, B.A. Ed. in *Methods in Molecular Biology* 67: Humana Press, Totowa (1997) and the publication entitled "PCR Methods and Applications" (1991, Cold Spring Harbor Laboratory Press). In each of these PCR procedures, PCR primers on either side of the nucleic acid sequences to be amplified are added to a suitably prepared nucleic acid sample along with dNTPs and a thermostable polymerase such as Taq polymerase, Pfu polymerase, or Vent polymerase. The nucleic acid in the sample is denatured and the PCR primers are specifically hybridized to complementary nucleic acid sequences in the sample. The hybridized primers are extended. Thereafter, another cycle of denaturation, hybridization, and extension is initiated. The cycles are repeated multiple times to produce an amplified fragment containing the nucleic acid sequence between the primer sites. PCR has further been described in several patents including US Patents 4,683,195, 4,683,202 and 4,965,188.

The identification of biallelic markers as described above allows the design of appropriate oligonucleotides, which can be used as primers to amplify DNA fragments comprising the biallelic markers of the present invention. Amplification can be performed using the primers initially used to discover new biallelic markers which are described herein or any set of primers allowing the amplification of a DNA fragment comprising a biallelic marker of the present invention. Primers can be prepared by any suitable method. As for example, direct chemical synthesis by a method such as the phosphodiester method of Narang S.A. et al. (*Methods Enzymol.* 68:90-98, 1979), the phosphodiester method of Brown E.L. et al. (*Methods Enzymol.* 68:109-151, 1979), the diethylphosphoramidite method of Beaucage et al. (*Tetrahedron Lett.* 22:1859-1862, 1981) and the solid support method described in EP 0 707 592.

In some embodiments the present invention provides primers for amplifying a DNA fragment containing one or more biallelic markers of the present invention. Preferred amplification primers are listed in Figure 8. It will be appreciated that the primers listed are merely exemplary and that any other

set of primers which produce amplification products containing one or more biallelic markers of the present invention.

The primers are selected to be substantially complementary to the different strands of each specific sequence to be amplified. The length of the primers of the present invention can range from 8 to 100 nucleotides, preferably from 8 to 50, 8 to 30 or more preferably 8 to 25 nucleotides. Shorter primers tend to lack specificity for a target nucleic acid sequence and generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. Longer primers are expensive to produce and can sometimes self-hybridize to form hairpin structures. The formation of stable hybrids depends on the melting temperature (Tm) of the DNA. The Tm depends on the length of the primer, the ionic strength of the solution and the G+C content. The higher the G+C content of the primer, the higher is the melting temperature because G:C pairs are held by three H bonds whereas A:T pairs have only two. The G+C content of the amplification primers of the present invention preferably ranges between 10 and 75 %, more preferably between 35 and 60 %, and most preferably between 40 and 55 %. The appropriate length for primers under a particular set of assay conditions may be empirically determined by one of skill in the art.

The spacing of the primers determines the length of the segment to be amplified. In the context of the present invention amplified segments carrying biallelic markers can range in size from at least about 25 bp to 35 kbp. Amplification fragments from 25-3000 bp are typical, fragments from 50-1000 bp are preferred and fragments from 100-600 bp are highly preferred. It will be appreciated that amplification primers for the biallelic markers may be any sequence which allow the specific amplification of any DNA fragment carrying the markers. Amplification primers may be labeled or immobilized on a solid support as described in I "Biallelic Markers and Polynucleotides Comprising Biallelic Markers."

III.C: Methods of Genotyping DNA Samples for Biallelic Markers

Any method known in the art can be used to identify the nucleotide present at a biallelic marker site. Since the biallelic marker allele to be detected has been identified and specified in the present invention, detection will prove simple for one of ordinary skill in the art by employing any of a number of techniques. Many genotyping methods require the previous amplification of the DNA region carrying the biallelic marker of interest. While the amplification of target or signal is often preferred at present, ultra sensitive detection methods which do not require amplification are also encompassed by the present genotyping methods. Methods well-known to those skilled in the art that can be used to detect biallelic polymorphisms include methods such as, conventional dot blot analyzes, single strand conformational polymorphism analysis (SSCP) described by Orita et al. (*Proc. Natl. Acad. Sci. U.S.A*

86:27776-2770, 1989), denaturing gradient gel electrophoresis (DGGE), heteroduplex analysis, mismatch cleavage detection, and other conventional techniques as described in Sheffield, V.C. et al. (*Proc. Natl. Acad. Sci. USA* 49:699-706, 1991), White et al. (*Genomics* 12:301-306, 1992), Grompe, M. et al. (*Proc. Natl. Acad. Sci. USA* 86:5855-5892, 1989) and Grompe, M. (*Nature Genetics* 5:111-117, 1993). Another method for determining the identity of the nucleotide present at a particular polymorphic site employs a specialized exonuclease-resistant nucleotide derivative as described in US Patent 4,656,127.

Preferred methods involve directly determining the identity of the nucleotide present at a biallelic marker site by sequencing assay, enzyme-based mismatch detection assay, or hybridization assay. The following is a description of some preferred methods. A highly preferred method is the microsequencing technique. The term "sequencing assay" is used herein to refer to polymerase extension of duplex primer/template complexes and includes both traditional sequencing and microsequencing.

1. Sequencing assays

The nucleotide present at a polymorphic site can be determined by sequencing methods. In a preferred embodiment, DNA samples are subjected to PCR amplification before sequencing as described above. DNA sequencing methods are described in IIC.

Preferably, the amplified DNA is subjected to automated dideoxy terminator sequencing reactions using a dye-primer cycle sequencing protocol. Sequence analysis allows the identification of the base present at the biallelic marker site.

2. Microsequencing assays

In microsequencing methods, a nucleotide at the polymorphic site that is unique to one of the alleles in a target DNA is detected by a single nucleotide primer extension reaction. This method involves appropriate microsequencing primers which, hybridize just upstream of a polymorphic base of interest in the target nucleic acid. A polymerase is used to specifically extend the 3' end of the primer with one single ddNTP (chain terminator) complementary to the selected nucleotide at the polymorphic site. Next the identity of the incorporated nucleotide is determined in any suitable way.

Typically, microsequencing reactions are carried out using fluorescent ddNTPs and the extended microsequencing primers are analyzed by electrophoresis on ABI 377 sequencing machines to determine the identity of the incorporated nucleotide as described in EP 412 883. Alternatively capillary electrophoresis can be used in order to process a higher number of assays simultaneously. An example

Printed: 29-12-2004 DESC PCT/IB IB0000184

of a typical microsequencing procedure that can be used in the context of the present invention is provided in Example 2.

Different approaches can be used to detect the nucleotide added to the microsequencing primer. A homogeneous phase detection method based on fluorescence resonance energy transfer has been described by Chen and Kwok (*Nucleic Acids Research* 25:347-353 1997) and Chen et al. (*Proc. Natl. Acad. Sci. USA* 94/20 10756-10761,1997). In this method amplified genomic DNA fragments containing polymorphic sites are incubated with a 5'-fluorescein–labeled primer in the presence of allelic dye-labeled dideoxyribonucleoside triphosphates and a modified Taq polymerase. The dye-labeled primer is extended one base by the dye-terminator specific for the allele present on the template. At the end of the genotyping reaction, the fluorescence intensities of the two dyes in the reaction mixture are analyzed directly without separation or purification. All these steps can be performed in the same tube and the fluorescence changes can be monitored in real time. Alternatively, the extended primer may be analyzed by MALDI-TOF Mass Spectrometry. The base at the polymorphic site is identified by the mass added onto the microsequencing primer (see Haff L.A. and Smirnov I.P., *Genome Research*, 7:378-388, 1997).

Microsequencing may be achieved by the established microsequencing method or by developments or derivatives thereof. Alternative methods include several solid-phase microsequencing techniques. The basic microsequencing protocol is the same as described previously, except that the method is conducted as a heterogeneous phase assay, in which the primer or the target molecule is immobilized or captured onto a solid support. To simplify the primer separation and the terminal nucleotide addition analysis, oligonucleotides are attached to solid supports or are modified in such ways that permit affinity separation as well as polymerase extension. The 5' ends and internal nucleotides of synthetic oligonucleotides can be modified in a number of different ways to permit different affinity separation approaches, e.g., biotinylation. If a single affinity group is used on the oligonucleotides, the oligonucleotides can be separated from the incorporated terminator reagent. This eliminates the need of physical or size separation. More than one oligonucleotide can be separated from the terminator reagent and analyzed simultaneously if more than one affinity group is used. This permits the analysis of several nucleic acid species or more nucleic acid sequence information per extension reaction. The affinity group need not be on the priming oligonucleotide but could alternatively be present on the template. For example, immobilization can be carried out via an interaction between biotinylated DNA and streptavidin-coated microtitration wells or avidin-coated polystyrene particles. In the same manner oligonucleotides or templates may be attached to a solid support in a high-density format. In such solid phase microsequencing reactions, incorporated ddNTPs can be radiolabeled (Syvänen, *Clinica Chimica Acta* 226:225-236, 1994) or linked to fluorescein

Printed: 29-12-2004 DESC IB0000184

(Livak and Hainer, *Human Mutation* 3:379-385,1994). The detection of radiolabeled ddNTPs can be achieved through scintillation-based techniques. The detection of fluorescein-linked ddNTPs can be based on the binding of antifluorescein antibody conjugated with alkaline phosphatase, followed by incubation with a chromogenic substrate (such as *p*-nitrophenyl phosphate). Other possible reporter-detection pairs include: ddNTP linked to dinitrophenyl (DNP) and anti-DNP alkaline phosphatase conjugate (Harju et al., *Clin. Chem.* 39/11 2282-2287, 1993) or biotinylated ddNTP and horseradish peroxidase-conjugated streptavidin with *o*-phenylenediamine as a substrate (WO 92/15712). As yet another alternative solid-phase microsequencing procedure, Nyren et al. (*Analytical Biochemistry* 208:171-175, 1993) described a method relying on the detection of DNA polymerase activity by an enzymatic luminometric inorganic pyrophosphate detection assay (ELIDA).

Pastinen et al. (*Genome research* 7:606-614, 1997) describe a method for multiplex detection of single nucleotide polymorphism in which the solid phase minisequencing principle is applied to an oligonucleotide array format. High-density arrays of DNA probes attached to a solid support (DNA chips) are further described in III.C.5.

In one aspect the present invention provides polynucleotides and methods to genotype one or more biallelic markers of the present invention by performing a microsequencing assay. Preferred microsequencing primers include those being featured Figure 7. It will be appreciated that the microsequencing primers listed in Figure 7 are merely exemplary and that, any primer having a 3' end immediately adjacent to a polymorphic nucleotide may be used. Similarly, it will be appreciated that microsequencing analysis may be performed for any biallelic marker or any combination of biallelic markers of the present invention. One aspect of the present invention is a solid support which includes one or more microsequencing primers listed in Figure 7, or fragments comprising at least 8, at least 12, at least 15, or at least 20 consecutive nucleotides thereof and having a 3' terminus immediately upstream of the corresponding biallelic marker, for determining the identity of a nucleotide at biallelic marker site.

3. Mismatch detection assays based on polymerases and ligases

In one aspect the present invention provides polynucleotides and methods to determine the allele of one or more biallelic markers of the present invention in a biological sample, by mismatch detection assays based on polymerases and/or ligases. These assays are based on the specificity of polymerases and ligases. Polymerization reactions places particularly stringent requirements on correct base pairing of the 3' end of the amplification primer and the joining of two oligonucleotides hybridized to a target DNA sequence is quite sensitive to mismatches close to the ligation site, especially at the 3' end. The terms "enzyme based mismatch detection assay" are used herein to refer to any method of

85 28-02-2000

determining the allele of a biallelic marker based on the specificity of ligases and polymerases. Preferred methods are described below. Methods, primers and various parameters to amplify DNA fragments comprising biallelic markers of the present invention are further described above in III.B.

Allele specific amplification

Discrimination between the two alleles of a biallelic marker can also be achieved by allele specific amplification, a selective strategy, whereby one of the alleles is amplified without amplification of the other allele. This is accomplished by placing a polymorphic base at the 3' end of one of the amplification primers. Because the extension forms from the 3'end of the primer, a mismatch at or near this position has an inhibitory effect on amplification. Therefore, under appropriate amplification conditions, these primers only direct amplification on their complementary allele. Designing the appropriate allele-specific primer and the corresponding assay conditions are well with the ordinary skill in the art.

Ligation/amplification based methods

The "Oligonucleotide Ligation Assay" (OLA) uses two oligonucleotides which are designed to be capable of hybridizing to abutting sequences of a single strand of a target molecules. One of the oligonucleotides is biotinylated, and the other is detectably labeled. If the precise complementary sequence is found in a target molecule, the oligonucleotides will hybridize such that their termini abut, and create a ligation substrate that can be captured and detected. OLA is capable of detecting biallelic markers and may be advantageously combined with PCR as described by Nickerson D.A. et al. (*Proc. Natl. Acad. Sci. U.S.A.* 87:8923-8927, 1990). In this method, PCR is used to achieve the exponential amplification of target DNA, which is then detected using OLA.

Other methods which are particularly suited for the detection of biallelic markers include LCR (ligase chain reaction), Gap LCR (GLCR) which are described above in III.B. As mentioned above LCR uses two pairs of probes to exponentially amplify a specific target. The sequences of each pair of oligonucleotides, is selected to permit the pair to hybridize to abutting sequences of the same strand of the target. Such hybridization forms a substrate for a template-dependant ligase. In accordance with the present invention, LCR can be performed with oligonucleotides having the proximal and distal sequences of the same strand of a biallelic marker site. In one embodiment, either oligonucleotide will be designed to include the biallelic marker site. In such an embodiment, the reaction conditions are selected such that the oligonucleotides can be ligated together only if the target molecule either contains or lacks the specific nucleotide(s) that is complementary to the biallelic marker on the oligonucleotide. In an alternative embodiment, the oligonucleotides will not include the biallelic marker, such that when

Printed: 29-12-2004 DESC PCT/IB0 IB0000184

they hybridize to the target molecule, a "gap" is created as described in WO 90/01069. This gap is then "filled" with complementary dNTPs (as mediated by DNA polymerase), or by an additional pair of oligonucleotides. Thus at the end of each cycle, each single strand has a complement capable of serving as a target during the next cycle and exponential allele-specific amplification of the desired sequence is obtained.

Ligase/Polymerase-mediated Genetic Bit Analysis™ is another method for determining the identity of a nucleotide at a preselected site in a nucleic acid molecule (WO 95/21271). This method involves the incorporation of a nucleoside triphosphate that is complementary to the nucleotide present at the preselected site onto the terminus of a primer molecule, and their subsequent ligation to a second oligonucleotide. The reaction is monitored by detecting a specific label attached to the reaction's solid phase or by detection in solution.

4. Hybridization assay methods

A preferred method of determining the identity of the nucleotide present at a biallelic marker site involves nucleic acid hybridization. The hybridization probes, which can be conveniently used in such reactions, preferably include the probes defined herein. Any hybridization assay may be used including Southern hybridization, Northern hybridization, dot blot hybridization and solid-phase hybridization (see Sambrook et al., Molecular Cloning – A Laboratory Manual, Second Edition, Cold Spring Harbor Press, N.Y., 1989).

Hybridization refers to the formation of a duplex structure by two single stranded nucleic acids due to complementary base pairing. Hybridization can occur between exactly complementary nucleic acid strands or between nucleic acid strands that contain minor regions of mismatch. Specific probes can be designed that hybridize to one form of a biallelic marker and not to the other and therefore are able to discriminate between different allelic forms. Allele-specific probes are often used in pairs, one member of a pair showing perfect match to a target sequence containing the original allele and the other showing a perfect match to the target sequence containing the alternative allele. Hybridization conditions should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles. Stringent, sequence specific hybridization conditions, under which a probe will hybridize only to the exactly complementary target sequence are well known in the art (Sambrook et al., Molecular Cloning – A Laboratory Manual, Second Edition, Cold Spring Harbor Press, N.Y., 1989). Stringent conditions are sequence dependent and will be different in different circumstances. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. By way of example and not limitation,

87 28-02-2000

procedures using conditions of high stringency are as follows: Prehybridization of filters containing DNA is carried out for 8 h to overnight at 65□C in buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 µg/ml denatured salmon sperm DNA. Filters are hybridized for 48 h at 65□C, the preferred hybridization temperature, in prehybridization mixture containing 100 µg/ml denatured salmon sperm DNA and 5-20 X 10^6 cpm of ^{32}P-labeled probe. Alternatively, the hybridization step can be performed at 65□C in the presence of SSC buffer, 1 x SSC corresponding to 0.15M NaCl and 0.05 M Na citrate. Subsequently, filter washes can be done at 37□C for 1 h in a solution containing 2X SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA, followed by a wash in 0.1X SSC at 50□C for 45 min. Alternatively, filter washes can be performed in a solution containing 2 x SSC and 0.1% SDS, or 0.5 x SSC and 0.1% SDS, or 0.1 x SSC and 0.1% SDS at 68□C for 15 minute intervals. Following the wash steps, the hybridized probes are detectable by autoradiography. By way of example and not limitation, procedures using conditions of intermediate stringency are as follows: Filters containing DNA are prehybridized, and then hybridized at a temperature of 60□C in the presence of a 5 x SSC buffer and labeled probe. Subsequently, filters washes are performed in a solution containing 2x SSC at 50□C and the hybridized probes are detectable by autoradiography. Other conditions of high and intermediate stringency which may be used are well known in the art and as cited in Sambrook et al. (Molecular Cloning - A Laboratory Manual, Second Edition, Cold Spring Harbor Press, N.Y., 1989) and Ausubel et al. (Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y., 1989).

Although such hybridizations can be performed in solution, it is preferred to employ a solid-phase hybridization assay. The target DNA comprising a biallelic marker of the present invention may be amplified prior to the hybridization reaction. The presence of a specific allele in the sample is determined by detecting the presence or the absence of stable hybrid duplexes formed between the probe and the target DNA. The detection of hybrid duplexes can be carried out by a number of methods. Various detection assay formats are well known which utilize detectable labels bound to either the target or the probe to enable detection of the hybrid duplexes. Typically, hybridization duplexes are separated from unhybridized nucleic acids and the labels bound to the duplexes are then detected. Those skilled in the art will recognize that wash steps may be employed to wash away excess target DNA or probe. Standard heterogeneous assay formats are suitable for detecting the hybrids using the labels present on the primers and probes.

Two recently developed assays allow hybridization-based allele discrimination with no need for separations or washes (see Landegren U. et al., *Genome Research*, 8:769-776,1998). The TaqMan assay takes advantage of the 5' nuclease activity of Taq DNA polymerase to digest a DNA probe annealed specifically to the accumulating amplification product. TaqMan probes are labeled with a

donor-acceptor dye pair that interacts via fluorescence energy transfer. Cleavage of the TaqMan probe by the advancing polymerase during amplification dissociates the donor dye from the quenching acceptor dye, greatly increasing the donor fluorescence. All reagents necessary to detect two allelic variants can be assembled at the beginning of the reaction and the results are monitored in real time (see Livak et al., *Nature Genetics*, 9:341-342, 1995). In an alternative homogeneous hybridization-based procedure, molecular beacons are used for allele discriminations. Molecular beacons are hairpin-shaped oligonucleotide probes that report the presence of specific nucleic acids in homogeneous solutions. When they bind to their targets they undergo a conformational reorganization that restores the fluorescence of an internally quenched fluorophore (Tyagi et al., *Nature Biotechnology*, 16:49-53, 1998).

The polynucleotides provided herein can be used in hybridization assays for the detection of biallelic marker alleles in biological samples. These probes are characterized in that they preferably comprise between 8 and 50 nucleotides, and in that they are sufficiently complementary to a sequence comprising a biallelic marker of the present invention to hybridize thereto and preferably sufficiently specific to be able to discriminate the targeted sequence for only one nucleotide variation. The GC content in the probes of the invention usually ranges between 10 and 75 %, preferably between 35 and 60 %, and more preferably between 40 and 55 %. The length of these probes can range from 10, 15, 20, or 30 to at least 100 nucleotides, preferably from 10 to 50, more preferably from 18 to 35 nucleotides. A particularly preferred probe is 25 nucleotides in length. Preferably the biallelic marker is within 4 nucleotides of the center of the polynucleotide probe. In particularly preferred probes the biallelic marker is at the center of said polynucleotide. Shorter probes may lack specificity for a target nucleic acid sequence and generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. Longer probes are expensive to produce and can sometimes self-hybridize to form hairpin structures. Methods for the synthesis of oligonucleotide probes have been described above and can be applied to the probes of the present invention.

Preferably the probes of the present invention are labeled or immobilized on a solid support. Labels and solid supports are further described in I. Detection probes are generally nucleic acid sequences or uncharged nucleic acid analogs such as, for example peptide nucleic acids which are disclosed in International Patent Application WO 92/20702, morpholino analogs which are described in U.S. Patents Numbered 5,185,444; 5,034,506 and 5,142,047. The probe may have to be rendered "non-extendable" in that additional dNTPs cannot be added to the probe. In and of themselves analogs usually are non-extendable and nucleic acid probes can be rendered non-extendable by modifying the 3' end of the probe such that the hydroxyl group is no longer capable of participating in elongation. For example, the 3' end of the probe can be functionalized with the capture or detection label to thereby

consume or otherwise block the hydroxyl group. Alternatively, the 3' hydroxyl group simply can be cleaved, replaced or modified, U.S. Patent Application Serial No. 07/049,061 filed April 19, 1993 describes modifications, which can be used to render a probe non-extendable.

The probes of the present invention are useful for a number of purposes. They can be used in Southern hybridization to genomic DNA or Northern hybridization to mRNA. The probes can also be used to detect PCR amplification products. By assaying the hybridization to an allele specific probe, one can detect the presence or absence of a biallelic marker allele in a given sample.

High-Throughput parallel hybridizations in array format are specifically encompassed within "hybridization assays" and are described below.

Hybridization to addressable arrays of oligonucleotides

Hybridization assays based on oligonucleotide arrays rely on the differences in hybridization stability of short oligonucleotides to perfectly matched and mismatched target sequence variants. Efficient access to polymorphism information is obtained through a basic structure comprising high-density arrays of oligonucleotide probes attached to a solid support (the chip) at selected positions. Each DNA chip can contain thousands to millions of individual synthetic DNA probes arranged in a grid-like pattern and miniaturized to the size of a dime.

The chip technology has already been applied with success in numerous cases. For example, the screening of mutations has been undertaken in the BRCA1 gene, in *S. cerevisiae* mutant strains, and in the protease gene of HIV-1 virus (Hacia et al., *Nature Genetics*, 14(4):441-447, 1996; Shoemaker et al., *Nature Genetics*, 14(4):450-456, 1996; Kozal et al., *Nature Medicine*, 2:753-759, 1996). Chips of various formats for use in detecting biallelic polymorphisms can be produced on a customized basis by Affymetrix (GeneChip™), Hyseq (HyChip and HyGnostics), and Protogene Laboratories.

In general, these methods employ arrays of oligonucleotide probes that are complementary to target nucleic acid sequence segments from an individual which, target sequences include a polymorphic marker. EP785280 describes a tiling strategy for the detection of single nucleotide polymorphisms. Briefly, arrays may generally be "tiled" for a large number of specific polymorphisms. By "tiling" is generally meant the synthesis of a defined set of oligonucleotide probes which is made up of a sequence complementary to the target sequence of interest, as well as preselected variations of that sequence, e.g., substitution of one or more given positions with one or more members of the basis set of monomers, i.e. nucleotides. Tiling strategies are further described in PCT application No. WO 95/11995. In a particular aspect, arrays are tiled for a number of specific, identified biallelic marker sequences. In particular the array is tiled to include a number of detection blocks, each detection block being specific for a specific biallelic marker or a set of biallelic markers. For example, a detection

block may be tiled to include a number of probes, which span the sequence segment that includes a specific polymorphism. To ensure probes that are complementary to each allele, the probes are synthesized in pairs differing at the biallelic marker. In addition to the probes differing at the polymorphic base, monosubstituted probes are also generally tiled within the detection block. These monosubstituted probes have bases at and up to a certain number of bases in either direction from the polymorphism, substituted with the remaining nucleotides (selected from A, T, G, C and U). Typically the probes in a tiled detection block will include substitutions of the sequence positions up to and including those that are 5 bases away from the biallelic marker. The monosubstituted probes provide internal controls for the tiled array, to distinguish actual hybridization from artefactual cross-hybridization. Upon completion of hybridization with the target sequence and washing of the array, the array is scanned to determine the position on the array to which the target sequence hybridizes. The hybridization data from the scanned array is then analyzed to identify which allele or alleles of the biallelic marker are present in the sample. Hybridization and scanning may be carried out as described in PCT application No. WO 92/10092 and WO 95/11995 and US patent No. 5,424,186.

Thus, in some embodiments, the chips may comprise an array of nucleic acid sequences of fragments of about 15 nucleotides in length. In further embodiments, the chip may comprise an array including at least one of the sequences selected from the group consisting of SEQ ID Nos. 1-654 except SEQ ID Nos. 419-424, 490, 531 and 540 and the sequences complementary thereto, or a fragment thereof at least about 8 consecutive nucleotides, preferably 10, 15, 20, more preferably 25, 30, 40, 47, or 50 consecutive nucleotides. In some embodiments, the chip may comprise an array of at least 2, 3, 4, 5, 6, 7, 8 or more of these polynucleotides of the invention. Solid supports and polynucleotides of the present invention attached to solid supports are further described in I. Biallelic Markers and Polynucleotides Comprising Biallelic Markers.

5) Integrated Systems

Another technique, which may be used to analyze polymorphisms, includes multicomponent integrated systems, which miniaturize and compartmentalize processes such as PCR and capillary electrophoresis reactions in a single functional device. An example of such technique is disclosed in US patent 5,589,136, which describes the integration of PCR amplification and capillary electrophoresis in chips.

Integrated systems can be envisaged mainly when microfluidic systems are used. These systems comprise a pattern of microchannels designed onto a glass, silicon, quartz, or plastic wafer included on a microchip. The movements of the samples are controlled by electric, electroosmotic or hydrostatic forces applied across different areas of the microchip. For genotyping biallelic markers, the

microfluidic system may integrate nucleic acid amplification, microsequencing, capillary electrophoresis and a detection method such as laser-induced fluorescence detection.

IV. Methods of Genetic Analysis Using the Biallelic Markers of the Present Invention

Different methods are available for the genetic analysis of complex traits (see Lander and Schork, *Science*, 265, 2037-2048, 1994). The search for disease-susceptibility genes is conducted using two main methods: the linkage approach in which evidence is sought for cosegregation between a locus and a putative trait locus using family studies, and the association approach in which evidence is sought for a statistically significant association between an allele and a trait or a trait causing allele (Khoury J. et al., *Fundamentals of Genetic Epidemiology, Oxford University Press, NY*, 1993). In general, the biallelic markers of the present invention find use in any method known in the art to demonstrate a statistically significant correlation between a genotype and a phenotype. The biallelic markers may be used in parametric and non-parametric linkage analysis methods. Preferably, the biallelic markers of the present invention are used to identify genes associated with detectable traits using association studies, an approach which does not require the use of affected families and which permits the identification of genes associated with complex and sporadic traits.

The genetic analysis using the biallelic markers of the present invention may be conducted on any scale. The whole set of biallelic markers of the present invention or any subset of biallelic markers of the present invention may be used. In some embodiments a subset of biallelic markers corresponding to one or several candidate genes of the present invention may be used. In other embodiments a subset of biallelic markers corresponding to candidate genes from a given pathway of arachidonic acid metabolism may be used. Such pathways include the cycloxygenase pathway and the lipoxygenase pathway. Alternatively, a subset of biallelic markers of the present invention localised on a specific chromosome segment may be used. Further, any set of genetic markers including a biallelic marker of the present invention may be used. A set of biallelic polymorphisms that, could be used as genetic markers in combination with the biallelic markers of the present invention, has been described in WO 98/20165. As mentioned above, it should be noted that the biallelic markers of the present invention may be included in any complete or partial genetic map of the human genome. These different uses are specifically contemplated in the present invention and claims.

IV.A. Linkage Analysis

Linkage analysis is based upon establishing a correlation between the transmission of genetic markers and that of a specific trait throughout generations within a family. Thus, the aim of linkage analysis is to detect marker loci that show cosegregation with a trait of interest in pedigrees.

Parametric methods

When data are available from successive generations there is the opportunity to study the degree of linkage between pairs of loci. Estimates of the recombination fraction enable loci to be ordered and placed onto a genetic map. With loci that are genetic markers, a genetic map can be established, and then the strength of linkage between markers and traits can be calculated and used to indicate the relative positions of markers and genes affecting those traits (Weir, B.S., *Genetic data Analysis II: Methods for Discrete population genetic Data, Sinauer Assoc., Inc., Sunderland, MA, USA,* 1996). The classical method for linkage analysis is the logarithm of odds (lod) score method (see Morton N.E., *Am.J. Hum.Genet.*, 7:277-318, 1955; Ott J., *Analysis of Human Genetic Linkage, John Hopkins University Press, Baltimore,* 1991). Calculation of lod scores requires specification of the mode of inheritance for the disease (parametric method). Generally, the length of the candidate region identified using linkage analysis is between 2 and 20Mb. Once a candidate region is identified as described above, analysis of recombinant individuals using additional markers allows further delineation of the candidate region. Linkage analysis studies have generally relied on the use of a maximum of 5,000 microsatellite markers, thus limiting the maximum theoretical attainable resolution of linkage analysis to about 600 kb on average.

Linkage analysis has been successfully applied to map simple genetic traits that show clear Mendelian inheritance patterns and which have a high penetrance (i.e., the ratio between the number of affected carriers of allele *a* and the total number of *a* carriers in the population). However, parametric linkage analysis suffers from a variety of drawbacks. First, it is limited by its reliance on the choice of a genetic model suitable for each studied trait. Furthermore, as already mentioned, the resolution attainable using linkage analysis is limited, and complementary studies are required to refine the analysis of the typical 2Mb to 20Mb regions initially identified through linkage analysis. In addition, parametric linkage analysis approaches have proven difficult when applied to complex genetic traits, such as those due to the combined action of multiple genes and/or environmental factors. It is very difficult to model these factors adequately in a lod score analysis. In such cases, too large an effort and cost are needed to recruit the adequate number of affected families required for applying linkage analysis to these situations, as recently discussed by Risch, N. and Merikangas, K. (*Science*, 273:1516-1517, 1996).

Non-parametric methods

The advantage of the so-called non-parametric methods for linkage analysis is that they do not require specification of the mode of inheritance for the disease, they tend to be more useful for the

Printed: 29-12-2004 DESC IB0000184

analysis of complex traits. In non-parametric methods, one tries to prove that the inheritance pattern of a chromosomal region is not consistent with random Mendelian segregation by showing that affected relatives inherit identical copies of the region more often than expected by chance. Affected relatives should show excess "allele sharing" even in the presence of incomplete penetrance and polygenic inheritance. In non-parametric linkage analysis the degree of agreement at a marker locus in two individuals can be measured either by the number of alleles identical by state (IBS) or by the number of alleles identical by descent (IBD). Affected sib pair analysis is a well-known special case and is the simplest form of these methods.

The biallelic markers of the present invention may be used in both parametric and non-parametric linkage analysis. Preferably biallelic markers may be used in non-parametric methods which allow the mapping of genes involved in complex traits. The biallelic markers of the present invention may be used in both IBD- and IBS- methods to map genes affecting a complex trait. In such studies, taking advantage of the high density of biallelic markers, several adjacent biallelic marker loci may be pooled to achieve the efficiency attained by multi-allelic markers (Zhao et al., *Am. J. Hum. Genet.*, 63:225-240, 1998).

However, both parametric and non-parametric linkage analysis methods analyse affected relatives, they tend to be of limited value in the genetic analysis of drug responses or in the analysis of side effects to treatments. This type of analysis is impractical in such cases due to the lack of availability of familial cases. In fact, the likelihood of having more than one individual in a family being exposed to the same drug at the same time is extremely low.

IV.B. Population Association Studies

The present invention comprises methods for identifying one or several genes among a set of candidate genes that are associated with a detectable trait using the biallelic markers of the present invention. In one embodiment the present invention comprises methods to detect an association between a biallelic marker allele or a biallelic marker haplotype and a trait. Further, the invention comprises methods to identify a trait causing allele in linkage disequilibrium with any biallelic marker allele of the present invention.

As described above, alternative approaches can be employed to perform association studies: genome-wide association studies, candidate region association studies and candidate gene association studies. In a preferred embodiment, the biallelic markers of the present invention are used to perform candidate gene association studies. The candidate gene analysis clearly provides a short-cut approach to the identification of genes and gene polymorphisms related to a particular trait when some information concerning the biology of the trait is available. Further, the biallelic markers of the present invention

28-02-2000

may be incorporated in any map of genetic markers of the human genome in order to perform genome-wide association studies. Methods to generate a high-density map of biallelic markers have been described in WIPO Patent application serial number PCT/IB98/01193. The biallelic markers of the present invention may further be incorporated in any map of a specific candidate region of the genome (a specific chromosome or a specific chromosomal segment for example).

As mentioned above, association studies may be conducted within the general population and are not limited to studies performed on related individuals in affected families. Association studies are extremely valuable as they permit the analysis of sporadic or multifactor traits. Moreover, association studies represent a powerful method for fine-scale mapping enabling much finer mapping of trait causing alleles than linkage studies. Studies based on pedigrees often only narrow the location of the trait causing allele. Association studies using the biallelic markers of the present invention can therefore be used to refine the location of a trait causing allele in a candidate region identified by Linkage Analysis methods. Moreover, once a chromosome segment of interest has been identified, the presence of a candidate gene such as a candidate gene of the present invention, in the region of interest can provide a shortcut to the identification of the trait causing allele. Biallelic markers of the present invention can be used to demonstrate that a candidate gene is associated with a trait. Such uses are specifically contemplated in the present invention and claims.

1. Determining the frequency of a biallelic marker allele or of a biallelic marker haplotype in a population

Association studies explore the relationships among frequencies for sets of alleles between loci. In addition, the present invention provides methods of determining the frequency in a population of an allele of a 12-LO- related biallelic marker comprising: a) genotyping individuals from said population for said biallelic marker and, b) determining the proportional representation of said biallelic marker in said population. Optionally, said 12-LO-related biallelic marker is selected from the biallelic markers described in Table I. The present invention further provides methods of estimating the frequency of a haplotype for a set of biallelic markers in a population, comprising: a) genotyping each individual in said population for at least one 12-LO-related biallelic marker; b) genotyping each individual in said population for a second biallelic marker by determining the identity of the nucleotides at said second biallelic marker for both copies of said second biallelic marker present in the genome; and c) applying a haplotype determination method to the identities of the nucleotides determined in steps a) and b) to obtain an estimate of said frequency. Optionally, said haplotype determination method is selected from asymmetric PCR amplification, double PCR amplification of specific alleles, the Clark method, or an

Printed: 29-12-2004 DESC IB0000184

expectation maximization algorithm. Optionally, said 12-LO-related biallelic marker is selected from the biallelic markers described in Table I.

Determining the frequency of an allele in a population

Allelic frequencies of the biallelic markers in a population can be determined using one of the methods described above under the heading "Methods for genotyping an individual for biallelic markers", or any genotyping procedure suitable for this intended purpose. Genotyping pooled samples or individual samples can determine the frequency of a biallelic marker allele in a population. One way to reduce the number of genotypings required is to use pooled samples. A major obstacle in using pooled samples is in terms of accuracy and reproducibility for determining accurate DNA concentrations in setting up the pools. Genotyping individual samples provides higher sensitivity, reproducibility and accuracy and; is the preferred method used in the present invention. Preferably, each individual is genotyped separately and simple gene counting is applied to determine the frequency of an allele of a biallelic marker or of a genotype in a given population.

Determining the frequency of a haplotype in a population

The gametic phase of haplotypes is unknown when diploid individuals are heterozygous at more than one locus. Using genealogical information in families gametic phase can sometimes be inferred (Perlin et al., *Am. J. Hum. Genet.*, 55:777-787, 1994). When no genealogical information is available different strategies may be used. One possibility is that the multiple-site heterozygous diploids can be eliminated from the analysis, keeping only the homozygotes and the single-site heterozygote individuals, but this approach might lead to a possible bias in the sample composition and the underestimation of low-frequency haplotypes. Another possibility is that single chromosomes can be studied independently, for example, by asymmetric PCR amplification (see Newton et al., *Nucleic Acids Res.*, 17:2503-2516, 1989; Wu et al., *Proc. Natl. Acad. Sci. USA*, 86:2757, 1989) or by isolation of single chromosome by limit dilution followed by PCR amplification (see Ruano et al., *Proc. Natl. Acad. Sci. USA*, 87:6296-6300, 1990). Further, a sample may be haplotyped for sufficiently close biallelic markers by double PCR amplification of specific alleles (Sarkar, G. and Sommer S.S., *Biotechniques*, 1991). These approaches are not entirely satisfying either because of their technical complexity, the additional cost they entail, their lack of generalisation at a large scale, or the possible biases they introduce. To overcome these difficulties, an algorithm to infer the phase of PCR-amplified DNA genotypes introduced by Clark A.G. (*Mol. Biol. Evol.*, 7:111-122, 1990) may be used. Briefly, the principle is to start filling a preliminary list of haplotypes present in the sample by examining unambiguous individuals, that is, the complete homozygotes and the single-site heterozygotes. Then

28-02-2000

other individuals in the same sample are screened for the possible occurrence of previously recognised haplotypes. For each positive identification, the complementary haplotype is added to the list of recognised haplotypes, until the phase information for all individuals is either resolved or identified as unresolved. This method assigns a single haplotype to each multiheterozygous individual, whereas several haplotypes are possible when there are more than one heterozygous site. Alternatively, one can use methods estimating haplotype frequencies in a population without assigning haplotypes to each individual. Preferably, a method based on an expectation-maximization (EM) algorithm (Dempster et al., *J. R. Stat. Soc.*, 39B: 1-38, 1977) leading to maximum-likelihood estimates of haplotype frequencies under the assumption of Hardy-Weinberg proportions (random mating) is used (see Excoffier L. and Slatkin M., *Mol. Biol. Evol.*, 12(5): 921-927, 1995). The EM algorithm is a generalised iterative maximum-likelihood approach to estimation that is useful when data are ambiguous and/or incomplete. The EM algorithm is used to resolve heterozygotes into haplotypes. Haplotype estimations are further described below under the heading "Statistical methods". Any other method known in the art to determine or to estimate the frequency of a haplotype in a population may also be used.

2. Linkage Disequilibrium analysis

Linkage disequilibrium is the non-random association of alleles at two or more loci and represents a powerful tool for mapping genes involved in disease traits (see Ajioka R.S. et al., *Am. J. Hum. Genet.*, 60:1439-1447, 1997). Biallelic markers, because they are densely spaced in the human genome and can be genotyped in more numerous numbers than other types of genetic markers (such as RFLP or VNTR markers), are particularly useful in genetic analysis based on linkage disequilibrium. The biallelic markers of the present invention may be used in any linkage disequilibrium analysis method known in the art.

When a disease mutation is first introduced into a population (by a new mutation or the immigration of a mutation carrier), it necessarily resides on a single chromosome and thus on a single "background" or "ancestral" haplotype of linked markers. Consequently, there is complete disequilibrium between these markers and the disease mutation: one finds the disease mutation only in the presence of a specific set of marker alleles. Through subsequent generations recombinations occur between the disease mutation and these marker polymorphisms, and the disequilibrium gradually dissipates. The pace of this dissipation is a function of the recombination frequency, so the markers closest to the disease gene will manifest higher levels of disequilibrium than those that are further away. When not broken up by recombination, "ancestral" haplotypes and linkage disequilibrium between marker alleles at different loci can be tracked not only through pedigrees but also through populations.

Linkage disequilibrium is usually seen as an association between one specific allele at one locus and another specific allele at a second locus.

The pattern or curve of disequilibrium between disease and marker loci is expected to exhibit a maximum that occurs at the disease locus. Consequently, the amount of linkage disequilibrium between a disease allele and closely linked genetic markers may yield valuable information regarding the location of the disease gene. For fine-scale mapping of a disease locus, it is useful to have some knowledge of the patterns of linkage disequilibrium that exist between markers in the studied region. As mentioned above the mapping resolution achieved through the analysis of linkage disequilibrium is much higher than that of linkage studies. The high density of biallelic markers combined with linkage disequilibrium analysis provides powerful tools for fine-scale mapping. Different methods to calculate linkage disequilibrium are described below under the heading "Statistical Methods".

3. Population-based case-control studies of trait-marker associations

As mentioned above, the occurrence of pairs of specific alleles at different loci on the same chromosome is not random and the deviation from random is called linkage disequilibrium. Association studies focus on population frequencies and rely on the phenomenon of linkage disequilibrium. If a specific allele in a given gene is directly involved in causing a particular trait, its frequency will be statistically increased in an affected (affected) population, when compared to the frequency in a trait negative population or in a random control population. As a consequence of the existence of linkage disequilibrium, the frequency of all other alleles present in the haplotype carrying the trait-causing allele will also be increased in affected (affected) individuals compared to trait negative individuals or random controls. Therefore, association between the trait and any allele (specifically a biallelic marker allele) in linkage disequilibrium with the trait-causing allele will suffice to suggest the presence of a trait-related gene in that particular region. Case-control populations can be genotyped for biallelic markers to identify associations that narrowly locate a trait causing allele. As any marker in linkage disequilibrium with one given marker associated with a trait will be associated with the trait. Linkage disequilibrium allows the relative frequencies in case-control populations of a limited number of genetic polymorphisms (specifically biallelic markers) to be analysed as an alternative to screening all possible functional polymorphisms in order to find trait-causing alleles. Association studies compare the frequency of marker alleles in unrelated case-control populations, and represent powerful tools for the dissection of complex traits.

Case-control populations (inclusion criteria)

Population-based association studies do not concern familial inheritance but compare the prevalence of a particular genetic marker, or a set of markers, in case-control populations. They are case-control studies based on comparison of unrelated case (affected or affected) individuals and unrelated control (unaffected or trait negative or random) individuals. Preferably the control group is composed of unaffected or trait negative individuals. Further, the control group is ethnically matched to the case population. Moreover, the control group is preferably matched to the case-population for the main known confusion factor for the trait under study (for example age-matched for an age-dependent trait). Ideally, individuals in the two samples are paired in such a way that they are expected to differ only in their disease status. In the following "affected population", "case population" and "affected population" are used interchangeably.

An important step in the dissection of complex traits using association studies is the choice of case-control populations (see Lander and Schork, *Science*, 265, 2037-2048, 1994). A major step in the choice of case-control populations is the clinical definition of a given trait or phenotype. Any genetic trait may be analysed by the association method proposed here by carefully selecting the individuals to be included in the affected and control phenotypic groups. Four criteria are often useful: clinical phenotype, age at onset, family history and severity. The selection procedure for continuous or quantitative traits (such as blood pressure for example) involves selecting individuals at opposite ends of the phenotype distribution of the trait under study, so as to include in these affected and control individuals with non-overlapping phenotypes. Preferably, case-control populations consist of phenotypically homogeneous populations. Affected and control populations consist of phenotypically uniform populations of individuals representing each between 1 and 98%, preferably between 1 and 80%, more preferably between 1 and 50%, and more preferably between 1 and 30%, most preferably between 1 and 20% of the total population under study, and selected among individuals exhibiting non-overlapping phenotypes. The clearer the difference between the two trait phenotypes, the greater the probability of detecting an association with biallelic markers. The selection of those drastically different but relatively uniform phenotypes enables efficient comparisons in association studies and the possible detection of marked differences at the genetic level, provided that the sample sizes of the populations under study are significant enough.

In preferred embodiments, a first group of between 50 and 300 affected individuals, preferably about 100 individuals, are recruited according to their phenotypes. A similar number of trait negative individuals are included in such studies.

In the present invention, typical examples of inclusion criteria include a disease involving arachidonic acid metabolism or the evaluation of the response to a drug acting on arachidonic acid metabolism or side effects to treatment with drugs acting on arachidonic acid metabolism.

Suitable examples of association studies using biallelic markers including the biallelic markers of the present invention, are studies involving the following populations:

a case population suffering from a disease involving arachidonic acid metabolism and a healthy unaffected control population, or

a case population treated with agents acting on arachidonic acid metabolism suffering from side-effects resulting from the treatment and a control population treated with the same agents showing no side-effects, or

a case population treated with agents acting on arachidonic acid metabolism showing a beneficial response and a control population treated with same agents showing no beneficial response.

In a preferred embodiment, eicosanoid related-markers may be used to identify individuals who are prone to hepatoxicity as a result of drug treatment. This includes diagnostic and prognostic assays to identify individuals who are prone to liver toxicity as a result of drug treatment, as well as clinical trials and treatment regimes which utilize these assays. Said drug treatment may include any pahrmaceutical compound suspected or known in the art to result in an increased level of hepatoxicity.

In another preferred embodiment, the trait considered was a side effect upon drug treatment; the study involved two populations derived from a clinical study of the anti-asthmatic drug zileuton. The case population was composed of asthmatic individuals treated with Zileuton showing zileuton-associated hepatotoxicity monitored by the serum level of alanine aminotransferase (ALT) and the control population was composed of asthmatic individuals treated with zileuton and having no increased serum level of ALT. Inclusion criteria and association between the biallelic markers of the present invention and zileuton-associated hepatotoxicity are further described below in IV.E. Association of Biallelic Markers of the Invention with Hepatoxicity to Anti-Asthma Drug Zileuton and in Example 5, Association between Side Effects upon Treatment with the Anti-Asthmatic Drug Zileuton (Zyflo™) and the Biallelic Markers of the 12-lipoxygenase Gene.

Association analysis

The general strategy to perform association studies using biallelic markers derived from a region carrying a candidate gene is to scan two groups of individuals (case-control populations) in order to measure and statistically compare the allele frequencies of the biallelic markers of the present invention in both groups.

If a statistically significant association with a trait is identified for at least one or more of the analysed biallelic markers, one can assume that: either the associated allele is directly responsible for causing the trait (the associated allele is the trait causing allele), or more likely the associated allele is in linkage disequilibrium with the trait causing allele. The specific characteristics of the associated allele

THIS PAGE BLANK (USPTO)

Printed: 29-12-2004 DESC PCT/IB00/00184 IB0000184

with respect to the candidate gene function usually gives further insight into the relationship between the associated allele and the trait (causal or in linkage disequilibrium). If the evidence indicates that the associated allele within the candidate gene is most probably not the trait causing allele but is in linkage disequilibrium with the real trait causing allele, then the trait causing allele can be found by sequencing the vicinity of the associated marker.

Association studies are usually run in two successive steps. In a first phase, the frequencies of a reduced number of biallelic markers from one or several candidate genes are determined in the affected and control populations. In a second phase of the analysis, the identity of the candidate gene and the position of the genetic loci responsible for the given trait is further refined using a higher density of markers from the relevant region. However, if the candidate gene under study is relatively small in length, as it is the case for many of the candidate genes analysed included in the present invention, a single phase may be sufficient to establish significant associations.

Haplotype analysis

As described above, when a chromosome carrying a disease allele first appears in a population as a result of either mutation or migration, the mutant allele necessarily resides on a chromosome having a set of linked markers: the ancestral haplotype. This haplotype can be tracked through populations and its statistical association with a given trait can be analysed. Complementing single point (allelic) association studies with multi-point association studies also called haplotype studies increases the statistical power of association studies. Thus, a haplotype association study allows one to define the frequency and the type of the ancestral carrier haplotype. A haplotype analysis is important in that it increases the statistical power of an analysis involving individual markers.

In a first stage of a haplotype frequency analysis, the frequency of the possible haplotypes based on various combinations of the identified biallelic markers of the invention is determined. The haplotype frequency is then compared for distinct populations of affected and control individuals. The number of affected individuals, which should be, subjected to this analysis to obtain statistically significant results usually ranges between 30 and 300, with a preferred number of individuals ranging between 50 and 150. The same considerations apply to the number of unaffected individuals (or random control) used in the study. The results of this first analysis provide haplotype frequencies in case-control populations, for each evaluated haplotype frequency a p-value and an odd ratio are calculated. If a statistically significant association is found the relative risk for an individual carrying the given haplotype of being affected with the trait under study can be approximated.

Interaction Analysis

The biallelic markers of the present invention may also be used to identify patterns of biallelic markers associated with detectable traits resulting from polygenic interactions. The analysis of genetic interaction between alleles at unlinked loci requires individual genotyping using the techniques described herein. The analysis of allelic interaction among a selected set of biallelic markers with appropriate level of statistical significance can be considered as a haplotype analysis. Interaction analysis consists in stratifying the case-control populations with respect to a given haplotype for the first loci and performing a haplotype analysis with the second loci with each subpopulation.

Statistical methods used in association studies are further described below in IV.C "Statistical Methods."

4. Testing for linkage in the presence of association

The biallelic markers of the present invention may further be used in TDT (transmission/disequilibrium test). TDT tests for both linkage and association and is not affected by population stratification. TDT requires data for affected individuals and their parents or data from unaffected sibs instead of from parents (see Spielmann S. et al., *Am. J. Hum. Genet.*, 52:506-516, 1993; Schaid D.J. et al., *Genet. Epidemiol.*,13:423-450, 1996, Spielmann S. and Ewens W.J., *Am. J. Hum. Genet.*, 62:450-458, 1998). Such combined tests generally reduce the false – positive errors produced by separate analyses.

IV.C. Statistical Methods

In general, any method known in the art to test whether a trait and a genotype show a statistically significant correlation may be used.

1. Methods in linkage analysis

Statistical methods and computer programs useful for linkage analysis are well-known to those skilled in the art (see Terwilliger J.D. and Ott J., *Handbook of Human Genetic Linkage, John Hopkins University Press, London,* 1994; Ott J., *Analysis of Human Genetic Linkage, John Hopkins University Press, Baltimore,* 1991).

2. Methods to estimate haplotype frequencies in a population

As described above, when genotypes are scored, it is often not possible to distinguish heterozygotes so that haplotype frequencies cannot be easily inferred. When the gametic phase is not known, haplotype frequencies can be estimated from the multilocus genotypic data. Any method

known to person skilled in the art can be used to estimate haplotype frequencies (see Lange K., *Mathematical and Statistical Methods for Genetic Analysis, Springer, New York*, 1997; Weir, B.S., *Genetic data Analysis II: Methods for Discrete population genetic Data, Sinauer Assoc., Inc., Sunderland, MA, USA*, 1996) Preferably, maximum-likelihood haplotype frequencies are computed using an Expectation- Maximization (EM) algorithm (see Dempster et al., *J. R. Stat. Soc.*, 39B:1-38, 1977; Excoffier L. and Slatkin M., *Mol. Biol. Evol.*, 12(5): 921-927, 1995). This procedure is an iterative process aiming at obtaining maximum-likelihood estimates of haplotype frequencies from multi-locus genotype data when the gametic phase is unknown. Haplotype estimations are usually performed by applying the EM algorithm using for example the EM-HAPLO program (Hawley M.E. et al., *Am. J. Phys. Anthropol.*, 18:104, 1994) or the Arlequin program (Schneider et al., *Arlequin: a software for population genetics data analysis*, University of Geneva, 1997). The EM algorithm is a generalised iterative maximum likelihood approach to estimation and is briefly described below.

In the following part of this text, phenotypes will refer to multi-locus genotypes with unknown phase. Genotypes will refer to known-phase multi-locus genotypes.

Suppose a sample of N unrelated individuals typed for K markers. The data observed are the unknown-phase K-locus phenotypes that can categorised in F different phenotypes. Suppose that we have H underlying possible haplotypes (in case of K biallelic markers, $H=2^K$).

For phenotype j, suppose that c_j genotypes are possible. We thus have the following equation

$$P_j = \sum_{i=1}^{c_j} pr(genotype_i) = \sum_{i=1}^{c_j} pr(h_k, h_l) \qquad \text{Equation 1}$$

where Pj is the probability of the phenotype j, h_k and h_l are the two haplotypes constituent the genotype i. Under the Hardy-Weinberg equilibrium, $pr(h_k,h_l)$ becomes :

$$pr(h_k, h_l) = pr(h_k)^2 \text{ if } h_k = h_l,\ pr(h_k, h_l) = 2\,pr(h_k).pr(h_l) \text{ if } h_k \neq h_l. \qquad \text{Equation 2}$$

The successive steps of the E-M algorithm can be described as follows:

Starting with initial values of the of haplotypes frequencies, noted $p_1^{(0)}, p_2^{(0)}, \ldots . p_H^{(0)}$, these initial values serve to estimate the genotype frequencies (Expectation step) and then estimate another set of haplotype frequencies (Maximisation step), noted $p_1^{(1)}, p_2^{(1)}, \ldots . p_H^{(1)}$, these two steps are iterated until changes in the sets of haplotypes frequency are very small.

A stop criterion can be that the maximum difference between haplotype frequencies between two iterations is less than 10^{-7}. These values can be adjusted according to the desired precision of estimations.

In details, at a given iteration s, the Expectation step consists in calculating the genotypes frequencies by the following equation:

$$pr(genotype_i)^{(s)} = pr(phenotype_j).pr(genotype_i|phenotype_j)^{(s)}$$

$$= \frac{n_j}{N}.\frac{pr(h_k,h_l)^{(s)}}{P_j^{(s)}} \qquad \text{Equation 3}$$

where genotype i occurs in phenotype j, and where h_k and h_l constitute genotype i. Each probability is derived according to eq.1, and eq.2 described above.

Then the Maximisation step simply estimates another set of haplotype frequencies given the genotypes frequencies. This approach is also known as gene-counting method (Smith, *Ann. Hum. Genet.*, 21:254-276, 1957).

$$p_t^{(s+1)} = \frac{1}{2}\sum_{j=1}^{F}\sum_{i=1}^{c_j}\delta_{it}.pr(genotype_i)^{(s)} \qquad \text{Equation 4}$$

Where δ_{it} is an indicator variable which count the number of time haplotype t in genotype i. It takes the values of 0, 1 or 2.

To ensure that the estimation finally obtained is the maximum-likelihood estimation several values of departures are required. The estimations obtained are compared and if they are different the estimations leading to the best likelihood are kept.

3. Methods to calculate linkage disequilibrium between markers

A number of methods can be used to calculate linkage disequilibrium between any two genetic positions, in practice linkage disequilibrium is measured by applying a statistical association test to haplotype data taken from a population.

Linkage disequilibrium between any pair of biallelic markers comprising at least one of the biallelic markers of the present invention (M_i, M_j) having alleles (a_i/b_i) at marker M_i and alleles (a_j/b_j) at marker M_j can be calculated for every allele combination (a_i,a_j ; a_i,b_j, b_i,a_j and b_i,b_j), according to the Piazza formula :

$\Delta_{aiaj} = \sqrt{\theta 4} - \sqrt{(\theta 4 + \theta 3)(\theta 4 + \theta 2)}$, where :

$\theta 4$= - - = frequency of genotypes not having allele a_i at M_i and not having allele a_j at M_j

$\theta 3$= - + = frequency of genotypes not having allele a_i at M_i and having allele a_j at M_j

$\theta 2$= + - = frequency of genotypes having allele a_i at M_i and not having allele a_j at M_j

Linkage disequilibrium (LD) between pairs of biallelic markers (M_i, M_j) can also be calculated for every allele combination (ai,aj, ai,bj ; b_i,a_j and b_i,b_j), according to the maximum-likelihood estimate (MLE) for delta (the composite genotypic disequilibrium coefficient), as described by Weir (Weir B.S., *Genetic Data Analysis, Sinauer Ass. Eds,* 1996). The MLE for the composite linkage disequilibrium is:

$D_{aiaj} = (2n_1 + n_2 + n_3 + n_4/2)/N - 2(pr(a_i).pr(a_j))$

Where $n_1 = \Sigma$ phenotype (a_i/a_i, a_j/a_j), $n_2 = \Sigma$ phenotype (a_i/a_i, a_j/b_j), $n_3 = \Sigma$ phenotype (a_i/b_i, a_j/a_j), n4= Σ phenotype (a_i/b_i, a_j/b_j) and N is the number of individuals in the sample.

This formula allows linkage disequilibrium between alleles to be estimated when only genotype, and not haplotype, data are available.

Another means of calculating the linkage disequilibrium between markers is as follows. For a couple of biallelic markers, M_i (a_i/b_i) and M_j (a_j/b_j), fitting the Hardy-Weinberg equilibrium, one can estimate the four possible haplotype frequencies in a given population according to the approach described above.

The estimation of gametic disequilibrium between *ai* and *aj* is simply:

$$D_{aiaj} = pr(haplotype(a_i, a_j)) - pr(a_i).pr(a_j).$$

Where $pr(a_i)$ is the probability of allele a_i and pr(a_j) is the probability of allele a_j and where $pr(haplotype\ (a_i, a_j))$ is estimated as in Equation 3 above.

For a couple of biallelic marker only one measure of disequilibrium is necessary to describe the association between M_i and M_j.

Then a normalised value of the above is calculated as follows:

$$D'_{aiaj} = D_{aiaj} / \max(-pr(a_i).pr(a_j), -pr(b_i).pr(b_j)) \text{ with } D_{aiaj}<0$$

$$D'_{aiaj} = D_{aiaj} / \max(pr(b_i).pr(a_j), pr(a_i).pr(b_j)) \text{ with } D_{aiaj}>0$$

The skilled person will readily appreciate that other LD calculation methods can be used without undue experimentation.

Linkage disequilibrium among a set of biallelic markers having an adequate heterozygosity rate can be determined by genotyping between 50 and 1000 unrelated individuals, preferably between 75 and 200, more preferably around 100.

4. Testing for association

Methods for determining the statistical significance of a correlation between a phenotype and a genotype, in this case an allele at a biallelic marker or a haplotype made up of such alleles, may be determined by any statistical test known in the art and with any accepted threshold of statistical significance being required. The application of particular methods and thresholds of significance are well with in the skill of the ordinary practitioner of the art.

Testing for association is performed by determining the frequency of a biallelic marker allele in case and control populations and comparing these frequencies with a statistical test to determine if their is a statistically significant difference in frequency which would indicate a correlation between the trait and the biallelic marker allele under study. Similarly, a haplotype analysis is performed by estimating

the frequencies of all possible haplotypes for a given set of biallelic markers in case and control populations, and comparing these frequencies with a statistical test to determine if their is a statistically significant correlation between the haplotype and the phenotype (trait) under study. Any statistical tool useful to test for a statistically significant association between a genotype and a phenotype may be used. Preferably the statistical test employed is a chi-square test with one degree of freedom. A P-value is calculated (the P-value is the probability that a statistic as large or larger than the observed one would occur by chance).

Statistical significance

In preferred embodiments, significance for diagnosis purposes, either as a positive basis for further diagnostic tests or as a preliminary starting point for early preventive therapy, the p value related to a biallelic marker association is preferably about 1 x 10-2 or less, more preferably about 1 x 10-4 or less, for a single biallelic marker analysis and about 1 x 10-3 or less, still more preferably 1 x 10-6 or less and most preferably of about 1 x 10-8 or less, for a haplotype analysis involving several markers. These values are believed to be applicable to any association studies involving single or multiple marker combinations.

The skilled person can use the range of values set forth above as a starting point in order to carry out association studies with biallelic markers of the present invention. In doing so, significant associations between the biallelic markers of the present invention and diseases involving arachidonic acid metabolism can be revealed and used for diagnosis and drug screening purposes.

Phenotypic permutation

In order to confirm the statistical significance of the first stage haplotype analysis described above, it might be suitable to perform further analyses in which genotyping data from case-control individuals are pooled and randomised with respect to the trait phenotype. Each individual genotyping data is randomly allocated to two groups, which contain the same number of individuals as the case-control populations used to compile the data obtained in the first stage. A second stage haplotype analysis is preferably run on these artificial groups, preferably for the markers included in the haplotype of the first stage analysis showing the highest relative risk coefficient. This experiment is reiterated preferably at least between 100 and 10000 times. The repeated iterations allow the determination of the percentage of obtained haplotypes with a significant p-value level.

Assessment of statistical association

To address the problem of false positives similar analysis may be performed with the same case-control populations in random genomic regions. Results in random regions and the candidate region are compared as described in US Provisional Patent Application entitled "Methods, software and apparati for identifying genomic regions harbouring a gene associated with a detectable trait".

5. Evaluation of risk factors

The association between a risk factor (in genetic epidemiology the risk factor is the presence or the absence of a certain allele or haplotype at marker loci) and a disease is measured by the odds ratio (OR) and by the relative risk (RR). If $P(R^+)$ is the probability of developing the disease for individuals with R and $P(R^-)$ is the probability for individuals without the risk factor, then the relative risk is simply the ratio of the two probabilities, that is:

$RR = P(R^+)/P(R^-)$

In case-control studies, direct measures of the relative risk cannot be obtained because of the sampling design. However, the odds ratio allows a good approximation of the relative risk for low-incidence diseases and can be calculated:

$$OR = \left[\frac{F^+}{1-F^+}\right] \Big/ \left[\frac{F^-}{(1-F^-)}\right]$$

F^+ is the frequency of the exposure to the risk factor in cases and F^- is the frequency of the exposure to the risk factor in controls. F^+ and F^- are calculated using the allelic or haplotype frequencies of the study and further depend on the underlying genetic model (dominant, recessive, additive...).

One can further estimate the attributable risk (AR) which describes the proportion of individuals in a population exhibiting a trait due to a given risk factor. This measure is important in quantitating the role of a specific factor in disease etiology and in terms of the public health impact of a risk factor. The public health relevance of this measure lies in estimating the proportion of cases of disease in the population that could be prevented if the exposure of interest were absent. AR is determined as follows:

$AR = P_E (RR-1) / (P_E (RR-1)+1)$

AR is the risk attributable to a biallelic marker allele or a biallelic marker haplotype. P_E is the frequency of exposure to an allele or a haplotype within the population at large; and RR is the relative risk which, is approximated with the odds ratio when the trait under study has a relatively low incidence in the general population.

IV.D. Association of Biallelic Markers of the Invention with Asthma

In the context of the present invention, an association between biallelic marker alleles from candidate genes of the present invention and a disease linked to arachidonic acid metabolism was demonstrated. The considered trait was asthma.

Asthma affects over 5% of the population in industrialized countries. It is increasing in prevalence and severity and has a rising mortality (Rang H.P., Ritter J.M. and Dale M.M.; *Pharmacology*; Churchill Livingstone, NY, 1995). Bronchial asthma is a multifactorial syndrome rather than a single disease, defined as airway obstruction characterized by inflammatory changes in the airways and bronchial hyper-responsiveness. In addition to the evidenced impact of environmental factors on the development of asthma, patterns of clustering and segregation in asthmatic families have suggested a genetic component to asthma. However the lack of a defined and specific asthma phenotype and of suitable markers for genetic analysis is proving to be a major hurdle for reliably identifying genes associated with asthma. The identification of genes implicated in asthma would represent a major step towards the identification of new molecular targets for the development of anti-asthma drugs. Moreover there is no straightforward physiological or biological blood test for the asthmatic state. As a result, adequate asthma treatment is often delayed, thereby allowing the inflammation process to better establish itself. Thus, there is a need for the identification of asthma susceptibility genes in order to develop an efficient and reliable asthma diagnostic test.

As mentioned above, products of arachidonic acid metabolism are important inflammatory mediators and have been involved in a number of inflammatory diseases including asthma. More specifically, prostaglandins and leukotrienes are thought to play a major role in the inflammatory process observed in asthma patients.

In order to investigate and identify a genetic origin to asthma a candidate gene scan for asthma was conducted. The rational of this approach was to: 1) select candidate genes potentially involved in the pathological pathway of interest, in this case arachidonic acid metabolism, 2) to identify biallelic markers in those genes and finally 3) to measure the frequency of biallelic marker alleles in order to determine if some alleles are more frequent in asthmatic populations than in non-affected populations. Results were further validated by haplotype studies. Significant associations between biallelic marker alleles from the FLAP and 12-LO genes and asthma were demonstrated in the context of the present invention. Association studies are further described in Examples 3 and 4.

This information is extremely valuable. The knowledge of a potential genetic predisposition, even if this predisposition is not absolute, might contribute in a very significant manner to treatment efficacy of asthma patients and to the development of diagnostic tools.

IV.E. Association of Biallelic Markers of the Invention with Hepatotoxicity to Anti-Asthma Drug Zileuton (Zyflo™)

In the context of the present invention, an association between the 12-LO gene and side effects related to treatment with the anti-asthmatic drug zileuton was discovered.

As mentioned above, bronchial asthma is a multifactorial syndrome rather than a single disease, defined as airway obstruction characterized by inflammatory changes in the airways and bronchial hyper-responsiveness. Although initially reversible with bronchiodilators, airway obstruction becomes increasingly irreversible if treated poorly. Asthma management therefore relies on early and regular use of drugs that control the disease. As a consequence, there is a strong need for efficient and safe therapeutic opportunities for patients with asthma. There are two main categories of anti-asthmatic drugs – bronchodilators and anti-inflammatory agents. There is now general agreement on the need to implement early anti-inflammatory treatment rather than relying on symptomatic treatment with bronchiodilators alone. The leukotrienes, a family of proinflammatory mediators arising via arachidonic acid metabolism, have been implicated in the inflammatory cascade that occurs in asthmatic airways. Of great relevance to the pathogenesis of asthma is the 5-lipoxygenase, which catalyzes the initial step in the biosynthesis of leukotrienes from arachidonic acid. Given the significant role of the inflammatory process in asthma, pharmacological agents, such as leukotriene antagonists and 5-lipoxygenase inhibitors have been developed.

Zileuton (Zyflo™) is an active inhibitor of 5-lipoxygenase, the enzyme that catalyzes the formation of leukotrienes from arachidonic acid, indicated for prophylaxis and chronic treatment of asthma. A minority of zileuton-treated patients develop liver function abnormalities. Close monitoring revealed that elevations of liver function tests may occur during treatment with zileuton. The ALT test (serum level of alanine aminotransferase) was used, which is considered the most sensitive indicator of liver injury.

In order to investigate and identify a genetic origin to zileuton-associated hepatotoxicity, a candidate gene scan was conducted. This approach comprised:

- selecting candidate genes potentially involved in the pathological pathway of interest or in the metabolism of zileuton, and
- identifying biallelic markers in those genes, and finally
- conducting association studies to identify biallelic marker alleles or haplotypes associated with elevations of liver function tests upon treatment with zileuton.

An association between elevated ALT levels upon treatment with zileuton and biallelic marker alleles from the 12-LO gene was demonstrated. Further details concerning this association study are provided in Example 5.

IV.F. Identification of Biallelic Markers in Linkage Disequilibrium with the Biallelic Markers of the Invention

Once a first biallelic marker has been identified in a genomic region of interest, the practitioner of ordinary skill in the art, using the teachings of the present invention, can easily identify additional biallelic markers in linkage disequilibrium with this first marker. As mentioned before any marker in linkage disequilibrium with a first marker associated with a trait will be associated with the trait. Therefore, once an association has been demonstrated between a given biallelic marker and a trait, the discovery of additional biallelic markers associated with this trait is of great interest in order to increase the density of biallelic markers in this particular region. The causal gene or mutation will be found in the vicinity of the marker or set of markers showing the highest correlation with the trait.

Identification of additional markers in linkage disequilibrium with a given marker involves: (a) amplifying a genomic fragment comprising a first biallelic marker from a plurality of individuals; (b) identifying of second biallelic markers in the genomic region harboring said first biallelic marker; (c) conducting a linkage disequilibrium analysis between said first biallelic marker and second biallelic markers; and (d) selecting said second biallelic markers as being in linkage disequilibrium with said first marker. Subcombinations comprising steps (b) and (c) are also contemplated.

Methods to identify biallelic markers and to conduct linkage disequilibrium analysis are described herein and can be carried out by the skilled person without undue experimentation. The present invention then also concerns biallelic markers which are in linkage disequilibrium with the specific biallelic markers shown in Figure 2 and which are expected to present similar characteristics in terms of their respective association with a given trait.

IV.G. Identification of Functional Mutations

Once a positive association is confirmed with a biallelic marker of the present invention, the associated candidate gene can be scanned for mutations by comparing the sequences of a selected number of affected individuals and control individuals. In a preferred embodiment, functional regions such as exons and splice sites, promoters and other regulatory regions of the candidate gene are scanned for mutations. Preferably, affected individuals carry the haplotype shown to be associated with the trait and trait negative or control individuals do not carry the haplotype or allele associated with the trait. The mutation detection procedure is essentially similar to that used for biallelic site identification.

The method used to detect such mutations generally comprises the following steps: (a) amplification of a region of the candidate gene comprising a biallelic marker or a group of biallelic markers associated with the trait from DNA samples of affected patients and trait negative controls; (b) sequencing of the amplified region; (c) comparison of DNA sequences from affected trait-positive patients and trait-negative controls; and (d) determination of mutations specific to affected trait-positive patients. Subcombinations which comprise steps (b) and (c) are specifically contemplated.

It is preferred that candidate polymorphisms be then verified by screening a larger population of cases and controls by means of any genotyping procedure such as those described herein, preferably using a microsequencing technique in an individual test format. Polymorphisms are considered as candidate mutations when present in cases and controls at frequencies compatible with the expected association results.

Identification of mutations and low frequency polymorphisms in the 5' flanking region of the 12-LO gene, in the exons and introns of the 12-LO gene and in the 3' flanking region of the 12-LO gene is further described in Example 5. Forty-nine low frequency polymorphisms and mutations were identified in the region of the 12-LO gene that was scanned. Low frequency polymorphisms and mutations identified in exons 5, 6, 8, and 13 are associated with amino acid substitutions at the polypeptide level. In each of these amino acid substitutions the original residue is replaced by a non-equivalent amino acid presenting different chemical properties. As a consequence, specificity, activity and function of the 12-LO enzyme are modified. Biallelic marker 10-343-231 is associated with a frame shift in the open reading frame of the 12-LO gene leading to the expression of a variant 12-LO polypeptide comprising only 131 amino acids. This mutant 12-LO enzyme is probably inactive or shows differences in specificity, activity and function. Biallelic marker 10-343-231 is associated with the deletion of a Leu residue in the 12-LO polypeptide.

Candidate polymorphisms and mutations of the 12-LO gene suspected of being responsible for the detectable phenotype, such as hepatoxicity to zileuton or asthma, can be confirmed by screening a larger population of affected and unaffected individuals using any of the genotyping procedures described herein. Preferably the microsequencing technique is used. Such polymorphisms are considered as candidate "trait-causing" mutations when they exhibit a statistically significant correlation with the detectable phenotype.

V. Biallelic Markers of the Invention in Methods of Genetic Diagnostics

The biallelic markers of the present invention can also be used to develop diagnostics tests capable of identifying individuals who express a detectable trait as the result of a specific genotype or individuals whose genotype places them at risk of developing a detectable trait at a subsequent time.

The trait analyzed using the present diagnostics may be any detectable trait, including a disease involving arachidonic acid metabolism, a response to an agent acting on arachidonic acid metabolism or side effects to an agent acting on arachidonic acid metabolism.

The diagnostic techniques of the present invention may employ a variety of methodologies to determine whether a test subject has a biallelic marker pattern associated with an increased risk of developing a detectable trait or whether the individual suffers from a detectable trait as a result of a particular mutation, including methods which enable the analysis of individual chromosomes for haplotyping, such as family studies, single sperm DNA analysis or somatic hybrids.

The present invention provides diagnostic methods to determine whether an individual is at risk of developing a disease or suffers from a disease resulting from a mutation or a polymorphism in a candidate gene of the present invention. The present invention also provides methods to determine whether an individual is likely to respond positively to an agent acting on arachidonic acid metabolism or whether an individual is at risk of developing an adverse side effect to an agent acting on arachidonic acid metabolism.

These methods involve obtaining a nucleic acid sample from the individual and, determining, whether the nucleic acid sample contains at least one allele or at least one biallelic marker haplotype, indicative of a risk of developing the trait or indicative that the individual expresses the trait as a result of possessing a particular candidate gene polymorphism or mutation (trait-causing allele).

Preferably, in such diagnostic methods, a nucleic acid sample is obtained from the individual and this sample is genotyped using methods described above in III. Methods of Genotyping an Individual for Biallelic Markers. The diagnostics may be based on a single biallelic marker or a on group of biallelic markers.

In each of these methods, a nucleic acid sample is obtained from the test subject and the biallelic marker pattern of one or more of the biallelic markers listed in Figure 2 is determined.

In one embodiment, a PCR amplification is conducted on the nucleic acid sample to amplify regions in which polymorphisms associated with a detectable phenotype have been identified. The amplification products are sequenced to determine whether the individual possesses one or more polymorphisms associated with a detectable phenotype. The primers used to generate amplification products may comprise the primers listed in Figure 8, or a preferred set of primers includes those described in SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; more preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652. Alternatively, the nucleic acid sample is subjected to microsequencing reactions as described above to determine whether the individual possesses one or more polymorphisms associated with a detectable phenotype resulting from a mutation or a polymorphism in a candidate gene. The primers used in the microsequencing reactions may

include the primers listed in Figure 7, or a preferred set of primers includes those described in SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; more preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652. In another embodiment, the nucleic acid sample is contacted with one or more allele specific oligonucleotide probes which, specifically hybridize to one or more candidate gene alleles associated with a detectable phenotype. The probes used in the hybridization assay may include the probes listed in Figure 9, or a preferred set of probes includes those described in SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; more preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652.

The present invention provides methods of determining whether an individual is at risk of developing asthma, or whether said individual suffers from asthma, comprising: a) genotyping said individual for at least one 12-LO-related biallelic marker; and b) correlating the result of step a) with a risk of developing asthma. In a preferred embodiment, said 12-LO-related biallelic marker is selected from the group consisting of biallelic markers: 12-197-244, 12-208-35, 12-226-167, 12-206-366, 10-346-141, 10-347-111, 10-347-165, 10-347-203, 10-347-220, 10-349-97, 10-349-224, 10-341-116, 12-196-119, 12-214-129, 12-216-421, 12-219-230, and 12-223-207. Preferably, said 12-LO-related biallelic marker is selected from the biallelic markers described in Example 5. The present invention also provides methods of determining whether an individual is at risk of developing hepatoxicity upon treatment with zileuton, comprising: a) genotyping said individual for at least one 12-LO-related biallelic marker; and b) correlating the result of step a) with a risk of developing hepatotoxicity upon treatment with zileuton. In a preferred embodiment, said 12-LO-related biallelic marker is selected from the group consisting of biallelic markers : 12-197-244, 12-208-35, 12-226-167, 12-206-366, 10-346-141, 10-347-111, 10-347-165, 10-347-203, 10-347-220, 10-349-97, 10-349-224, 10-341-116, 12-196-119, 12-214-129, 12-216-421, 12-219-230, and 12-223-207. Preferably, said 12-LO-related biallelic marker is selected from the biallelic markers described in Example 5, Association between Side Effects upon Treatment with the Anti-Asthmatic Drug Zileuton (Zyflo™) and the Biallelic Markers of the 12-lipoxygenase Gene.

These diagnostic methods are extremely valuable as they can, in certain circumstances, be used to initiate preventive treatments or to allow an individual carrying a significant haplotype to foresee warning signs such as minor symptoms. In diseases in which attacks may be extremely violent and sometimes fatal if not treated on time, such as asthma, the knowledge of a potential predisposition, even if this predisposition is not absolute, might contribute in a very significant manner to treatment efficacy. Similarly, a diagnosed predisposition to a potential side effect could immediately direct the physician toward a treatment for which such side effects have not been observed during clinical trials.

Diagnostics, which analyze and predict response to a drug or side effects to a drug, may be used to determine whether an individual should be treated with a particular drug. For example, if the diagnostic indicates a likelihood that an individual will respond positively to treatment with a particular drug, the drug may be administered to the individual. Conversely, if the diagnostic indicates that an individual is likely to respond negatively to treatment with a particular drug, an alternative course of treatment may be prescribed. A negative response may be defined as either the absence of an efficacious response or the presence of toxic side effects.

Clinical drug trials represent another application for the markers of the present invention. One or more markers indicative of response to an agent acting on arachidonic acid metabolism or to side effects to an agent acting on arachidonic acid metabolism may be identified using the methods described above. Thereafter, potential participants in clinical trials of such an agent may be screened to identify those individuals most likely to respond favorably to the drug and exclude those likely to experience side effects. In that way, the effectiveness of drug treatment may be measured in individuals who respond positively to the drug, without lowering the measurement as a result of the inclusion of individuals who are unlikely to respond positively in the study and without risking undesirable safety problems.

VI. Computer-Related Embodiments

As used herein the term "nucleic acid codes of the invention" encompass the nucleotide sequences comprising, consisting essentially of, or consisting of any one of the following: a) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500 or 1000 nucleotides, to the extent that a polynucleotide of these lengths is consistent with the lengths of the particular Sequence ID, of a sequence selected from the group consisting of the sequences described in Figure 3, and the complements thereof, excluding Sequence ID Nos. 1-10, 19, 23-25, and 647-650; b) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500 or 1000 nucleotides, to the extent that a polynucleotide of these lengths is consistent with the lengths of the particular Sequence ID, of a sequence selected from the group consisting of the sequences described in Figure 4, and the complements thereof, excluding Sequence ID Nos. 11-18 and 20-21; c) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, or 500 nucleotides, to the extent that a polynucleotide of these lengths is consistent with the lengths of the particular Sequence ID, of a sequence selected from the group consisting of the sequences described in Figure 7, more preferably a set of markers or sequences consisting of those markers or sequences found in SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652, and the complements thereof, wherein said span includes an eicosanoid-related biallelic marker, preferably an eicosanoid-related biallelic

marker described in Figure 2, preferably the biallelic markers found in SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; or more preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652, in said sequence with the alternative allele present at said biallelic marker.

The "nucleic acid codes of the invention" further encompass nucleotide sequences homologous to a contiguous span of at least 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500 or 1000 nucleotides, to the extent that a contiguous span of these lengths is consistent with the lengths of the particular Sequence ID, of a sequence selected from the group consisting of the sequences described in Figure 3, Figure 4 and Figure 7 and the complements thereof. Homologous sequences refer to a sequence having at least 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, or 75% homology to these contiguous spans. Homology may be determined using any method described herein, including BLAST2N with the default parameters or with any modified parameters. Homologous sequences also may include RNA sequences in which uridines replace the thymines in the nucleic acid codes of the invention. It will be appreciated that the nucleic acid codes of the invention can be represented in the traditional single character format (See the inside back cover of Stryer, Lubert. *Biochemistry*, 3rd edition. W. H Freeman & Co., New York.) or in any other format or code which records the identity of the nucleotides in a sequence.

It will be appreciated by those skilled in the art that the nucleic acid codes of the invention, one or more of the polypeptide codes of SEQ ID Nos. 653 and 654 can be stored, recorded, and manipulated on any medium which can be read and accessed by a computer. As used herein, the words "recorded" and "stored" refer to a process for storing information on a computer medium. A skilled artisan can readily adopt any of the presently known methods for recording information on a computer readable medium to generate manufactures comprising one or more of the nucleic acid codes of the invention and one or more of the polypeptide codes of SEQ ID Nos. 653-654. Another aspect of the present invention is a computer readable medium having recorded thereon at least 2, 5, 10, 15, 20, 25, 30, or 50 nucleic acid codes of the invention, and the complements thereof. Another aspect of the present invention is a computer readable medium having recorded thereon at least 2, 5, 10, 15, 20, 25, 30, or 50 polypeptide codes of SEQ ID Nos. 653-654.

Computer readable media include magnetically readable media, optically readable media, electronically readable media and magnetic/optical media. For example, the computer readable media may be a hard disk, a floppy disk, a magnetic tape, CD-ROM, Digital Versatile Disk (DVD), Random Access Memory (RAM), or Read Only Memory (ROM) as well as other types of other media known to those skilled in the art.

Embodiments of the present invention include systems, particularly computer systems which store and manipulate the sequence information described herein. One example of a computer system 100 is illustrated in block diagram form in Figure 22. As used herein, "a computer system" refers to the hardware

components, software components, and data storage components used to analyze the nucleotide sequences of the nucleic acid codes of the invention , or the amino acid sequences of the polypeptide codes of SEQ ID Nos. 653-654. In one embodiment, the computer system 100 is a Sun Enterprise 1000 server (Sun Microsystems, Palo Alto, CA). The computer system 100 preferably includes a processor for processing, accessing and manipulating the sequence data. The processor 105 can be any well-known type of central processing unit, such as the Pentium III from Intel Corporation, or similar processor from Sun, Motorola, Compaq or International Business Machines. Preferably, the computer system 100 is a general purpose system that comprises the processor 105 and one or more internal data storage components 110 for storing data, and one or more data retrieving devices for retrieving the data stored on the data storage components. A skilled artisan can readily appreciate that any one of the currently available computer systems are suitable. In one particular embodiment, the computer system 100 includes a processor 105 connected to a bus which is connected to a main memory 115 (preferably implemented as RAM) and one or more internal data storage devices 110, such as a hard drive and/or other computer readable media having data recorded thereon. In some embodiments, the computer system 100 further includes one or more data retrieving device 118 for reading the data stored on the internal data storage devices 110. The data retrieving device 118 may represent, for example, a floppy disk drive, a compact disk drive, a magnetic tape drive, etc. In some embodiments, the internal data storage device 110 is a removable computer readable medium such as a floppy disk, a compact disk, a magnetic tape, etc. containing control logic and/or data recorded thereon. The computer system 100 may advantageously include or be programmed by appropriate software for reading the control logic and/or the data from the data storage component once inserted in the data retrieving device. The computer system 100 includes a display 120 which is used to display output to a computer user. It should also be noted that the computer system 100 can be linked to other computer systems 125a-c in a network or wide area network to provide centralized access to the computer system 100. Software for accessing and processing the nucleotide sequences of the nucleic acid codes of the invention, or the amino acid sequences of the polypeptide codes of SEQ ID Nos. 653-654 (such as search tools, compare tools, and modeling tools etc.) may reside in main memory 115 during execution. In some embodiments, the computer system 100 may further comprise a sequence comparer for comparing the above-described nucleic acid codes of the invention or polypeptide codes of SEQ ID Nos. 653-654 stored on a computer readable medium to reference nucleotide or polypeptide sequences stored on a computer readable medium. A "sequence comparer" refers to one or more programs which are implemented on the computer system 100 to compare a nucleotide or polypeptide sequence with other nucleotide or polypeptide sequences and/or compounds including but not limited to peptides, peptidomimetics, and chemicals stored within the data storage means. For example, the sequence comparer may compare the nucleotide sequences of the nucleic acid codes of the invention, or the amino acid sequences of the

polypeptide codes of SEQ ID Nos. 653-654 stored on a computer readable medium to reference sequences stored on a computer readable medium to identify homologies, motifs implicated in biological function, or structural motifs. The various sequence comparer programs identified elsewhere in this patent specification are particularly contemplated for use in this aspect of the invention.

Figure 23 is a flow diagram illustrating one embodiment of a process 200 for comparing a new nucleotide or protein sequence with a database of sequences in order to determine the homology levels between the new sequence and the sequences in the database. The database of sequences can be a private database stored within the computer system 100, or a public database such as GENBANK, PIR OR SWISSPROT that is available through the Internet.

The process 200 begins at a start state 201 and then moves to a state 202 wherein the new sequence to be compared is stored to a memory in a computer system 100. As discussed above, the memory could be any type of memory, including RAM or an internal storage device. The process 200 then moves to a state 204 wherein a database of sequences is opened for analysis and comparison. The process 200 then moves to a state 206 wherein the first sequence stored in the database is read into a memory on the computer. A comparison is then performed at a state 210 to determine if the first sequence is the same as the second sequence. It is important to note that this step is not limited to performing an exact comparison between the new sequence and the first sequence in the database. Well-known methods are known to those of skill in the art for comparing two nucleotide or protein sequences, even if they are not identical. For example, gaps can be introduced into one sequence in order to raise the homology level between the two tested sequences. The parameters that control whether gaps or other features are introduced into a sequence during comparison are normally entered by the user of the computer system. Once a comparison of the two sequences has been performed at the state 210, a determination is made at a decision state 210 whether the two sequences are the same. Of course, the term "same" is not limited to sequences that are absolutely identical. Sequences that are within the homology parameters entered by the user will be marked as "same" in the process 200. If a determination is made that the two sequences are the same, the process 200 moves to a state 214 wherein the name of the sequence from the database is displayed to the user. This state notifies the user that the sequence with the displayed name fulfills the homology constraints that were entered. Once the name of the stored sequence is displayed to the user, the process 200 moves to a decision state 218 wherein a determination is made whether more sequences exist in the database. If no more sequences exist in the database, then the process 200 terminates at an end state 220. However, if more sequences do exist in the database, then the process 200 moves to a state 224 wherein a pointer is moved to the next sequence in the database so that it can be compared to the new sequence. In this manner, the new sequence is aligned and compared with every sequence in the database. It should be noted that if a determination had been made at the decision state 212 that the sequences were

not homologous, then the process 200 would move immediately to the decision state 218 in order to determine if any other sequences were available in the database for comparison. Accordingly, one aspect of the present invention is a computer system comprising a processor, a data storage device having stored thereon a nucleic acid code of the invention or a polypeptide code of SEQ ID Nos. 653-654, a data storage device having retrievably stored thereon reference nucleotide sequences or polypeptide sequences to be compared to the nucleic acid code of the invention or polypeptide code of SEQ ID Nos. 653-654 and a sequence comparer for conducting the comparison. The sequence comparer may indicate a homology level between the sequences compared or identify structural motifs in the above described nucleic acid code of the invention and polypeptide codes of SEQ ID Nos. 653-654 or it may identify structural motifs in sequences which are compared to these nucleic acid codes and polypeptide codes. In some embodiments, the data storage device may have stored thereon the sequences of at least 2, 5, 10, 15, 20, 25, 30, or 50 of the nucleic acid codes of the invention or polypeptide codes of SEQ ID Nos. 653-654.

Another aspect of the present invention is a method for determining the level of homology between a nucleic acid code of the invention and a reference nucleotide sequence, comprising the steps of reading the nucleic acid code and the reference nucleotide sequence through the use of a computer program which determines homology levels and determining homology between the nucleic acid code and the reference nucleotide sequence with the computer program. The computer program may be any of a number of computer programs for determining homology levels, including those specifically enumerated herein, including BLAST2N with the default parameters or with any modified parameters. The method may be implemented using the computer systems described above. The method may also be performed by reading 2, 5, 10, 15, 20, 25, 30, or 50 of the above described nucleic acid codes of the invention through use of the computer program and determining homology between the nucleic acid codes and reference nucleotide sequences .

Figure 24 is a flow diagram illustrating one embodiment of a process 250 in a computer for determining whether two sequences are homologous. The process 250 begins at a start state 252 and then moves to a state 254 wherein a first sequence to be compared is stored to a memory. The second sequence to be compared is then stored to a memory at a state 256. The process 250 then moves to a state 260 wherein the first character in the first sequence is read and then to a state 262 wherein the first character of the second sequence is read. It should be understood that if the sequence is a nucleotide sequence, then the character would normally be either A, T, C, G or U. If the sequence is a protein sequence, then it should be in the single letter amino acid code so that the first and sequence sequences can be easily compared.

A determination is then made at a decision state 264 whether the two characters are the same. If they are the same, then the process 250 moves to a state 268 wherein the next characters in the first and second sequences are read. A determination is then made whether the next characters are the same. If they are, then the process 250 continues this loop until two characters are not the same. If a determination is made that the next two characters are not the same, the process 250 moves to a decision state 274 to determine whether there are any more characters either sequence to read. If there aren't any more characters to read, then the process 250 moves to a state 276 wherein the level of homology between the first and second sequences is displayed to the user. The level of homology is determined by calculating the proportion of characters between the sequences that were the same out of the total number of sequences in the first sequence. Thus, if every character in a first 100 nucleotide sequence aligned with a every character in a second sequence, the homology level would be 100%. Alternatively, the computer program may be a computer program which compares the nucleotide sequences of the nucleic acid codes of the present invention, to reference nucleotide sequences in order to determine whether the nucleic acid code of SEQ ID Nos. 1-652 differs from a reference nucleic acid sequence at one or more positions. Optionally such a program records the length and identity of inserted, deleted or substituted nucleotides with respect to the sequence of either the reference polynucleotide or the nucleic acid code of SEQ ID Nos. 1-652. In one embodiment, the computer program may be a program which determines whether the nucleotide sequences of the nucleic acid codes of the invention contain a biallelic marker or single nucleotide polymorphism (SNP) with respect to a reference nucleotide sequence. This single nucleotide polymorphism may comprise a single base substitution, insertion, or deletion, while this biallelic marker may comprise abour one to ten consecutive bases substituted, inserted or deleted.

Another aspect of the present invention is a method for determining the level of homology between a polypeptide code of SEQ ID Nos. 653-654 and a reference polypeptide sequence, comprising the steps of reading the polypeptide code of SEQ ID Nos. 653-654 and the reference polypeptide sequence through use of a computer program which determines homology levels and determining homology between the polypeptide code and the reference polypeptide sequence using the computer program.

Accordingly, another aspect of the present invention is a method for determining whether a nucleic acid code of the invention differs at one or more nucleotides from a reference nucleotide sequence comprising the steps of reading the nucleic acid code and the reference nucleotide sequence through use of a computer program which identifies differences between nucleic acid sequences and identifying differences between the nucleic acid code and the reference nucleotide sequence with the computer program. In some embodiments, the computer program is a program which identifies single

nucleotide polymorphisms. The method may be implemented by the computer systems described above and the method illustrated in Figure 24. The method may also be performed by reading at least 2, 5, 10, 15, 20, 25, 30, or 50 of the nucleic acid codes of the invention and the reference nucleotide sequences through the use of the computer program and identifying differences between the nucleic acid codes and the reference nucleotide sequences with the computer program. In other embodiments the computer based system may further comprise an identifier for identifying features within the nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of SEQ ID Nos. 653-654. An "identifier" refers to one or more programs which identifies certain features within the above-described nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of SEQ ID Nos. 653-654. In one embodiment, the identifier may comprise a program which identifies an open reading frame in the cDNAs codes of SEQ ID No. 652.

Figure 25 is a flow diagram illustrating one embodiment of an identifier process 300 for detecting the presence of a feature in a sequence. The process 300 begins at a start state 302 and then moves to a state 304 wherein a first sequence that is to be checked for features is stored to a memory 115 in the computer system 100. The process 300 then moves to a state 306 wherein a database of sequence features is opened. Such a database would include a list of each feature's attributes along with the name of the feature. For example, a feature name could be "Initiation Codon" and the attribute would be "ATG." Another example would be the feature name "TAATAA Box" and the feature attribute would be "TAATAA". An example of such a database is produced by the University of Wisconsin Genetics Computer Group (www.gcg.com). Once the database of features is opened at the state 306, the process 300 moves to a state 308 wherein the first feature is read from the database. A comparison of the attribute of the first feature with the first sequence is then made at a state 310. A determination is then made at a decision state 316 whether the attribute of the feature was found in the first sequence. If the attribute was found, then the process 300 moves to a state 318 wherein the name of the found feature is displayed to the user. The process 300 then moves to a decision state 320 wherein a determination is made whether move features exist in the database. If no more features do exist, then the process 300 terminates at an end state 324. However, if more features do exist in the database, then the process 300 reads the next sequence feature at a state 326 and loops back to the state 310 wherein the attribute of the next feature is compared against the first sequence. It should be noted, that if the feature attribute is not found in the first sequence at the decision state 316, the process 300 moves directly to the decision state 320 in order to determine if any more features exist in the database. In another embodiment, the identifier may comprise a molecular modeling program which determines the 3-dimensional structure of the polypeptides codes of SEQ ID Nos. 653-654. In some embodiments,

the molecular modeling program identifies target sequences that are most compatible with profiles representing the structural environments of the residues in known three-dimensional protein structures. (See, e.g., Eisenberg et al., U.S. Patent No. 5,436,850 issued July 25, 1995). In another technique, the known three-dimensional structures of proteins in a given family are superimposed to define the structurally conserved regions in that family. This protein modeling technique also uses the known three-dimensional structure of a homologous protein to approximate the structure of the polypeptide codes of SEQ ID Nos. 653-654. (See e.g., Srinivasan, et al., U.S. Patent No. 5,557,535 issued September 17, 1996). Conventional homology modeling techniques have been used routinely to build models of proteases and antibodies. (Sowdhamini et al., Protein Engineering 10:207, 215 (1997)). Comparative approaches can also be used to develop three-dimensional protein models when the protein of interest has poor sequence identity to template proteins. In some cases, proteins fold into similar three-dimensional structures despite having very weak sequence identities. For example, the three-dimensional structures of a number of helical cytokines fold in similar three-dimensional topology in spite of weak sequence homology. The recent development of threading methods now enables the identification of likely folding patterns in a number of situations where the structural relatedness between target and template(s) is not detectable at the sequence level. Hybrid methods, in which fold recognition is performed using Multiple Sequence Threading (MST), structural equivalencies are deduced from the threading output using a distance geometry program DRAGON to construct a low resolution model, and a full-atom representation is constructed using a molecular modeling package such as QUANTA.

According to this 3-step approach, candidate templates are first identified by using the novel fold recognition algorithm MST, which is capable of performing simultaneous threading of multiple aligned sequences onto one or more 3-D structures. In a second step, the structural equivalencies obtained from the MST output are converted into interresidue distance restraints and fed into the distance geometry program DRAGON, together with auxiliary information obtained from secondary structure predictions. The program combines the restraints in an unbiased manner and rapidly generates a large number of low resolution model confirmations. In a third step, these low resolution model confirmations are converted into full-atom models and subjected to energy minimization using the molecular modeling package QUANTA. (See e.g., Aszódi et al., Proteins: Structure, Function, and Genetics, Supplement 1:38-42 (1997)).

The results of the molecular modeling analysis may then be used in rational drug design techniques to identify agents which modulate the activity of the polypeptide codes of SEQ ID Nos. 653-654. Accordingly, another aspect of the present invention is a method of identifying a feature within the nucleic acid codes of the invention or the polypeptide codes of SEQ ID Nos. 653-654 comprising

reading the nucleic acid code(s) or the polypeptide code(s) through the use of a computer program which identifies features therein and identifying features within the nucleic acid code(s) or polypeptide code(s) with the computer program. In one embodiment, computer program comprises a computer program which identifies open reading frames. In a further embodiment, the computer program identifies structural motifs in a polypeptide sequence. In another embodiment, the computer program comprises a molecular modeling program. The method may be performed by reading a single sequence or at least 2, 5, 10, 15, 20, 25, 30, or 50 of the nucleic acid codes of the invention or the polypeptide codes of SEQ ID Nos. 653-654 through the use of the computer program and identifying features within the nucleic acid codes or polypeptide codes with the computer program. The nucleic acid codes of the invention or the polypeptide codes of SEQ ID Nos. 653-654 may be stored and manipulated in a variety of data processor programs in a variety of formats. For example, the nucleic acid codes of the invention or the polypeptide codes of SEQ ID Nos. 653-654 may be stored as text in a word processing file, such as MicrosoftWORD or WORDPERFECT or as an ASCII file in a variety of database programs familiar to those of skill in the art, such as DB2, SYBASE, or ORACLE. In addition, many computer programs and databases may be used as sequence comparers, identifiers, or sources of reference nucleotide or polypeptide sequences to be compared to the nucleic acid codes of the invention or the polypeptide codes of SEQ ID Nos. 653-654. The following list is intended not to limit the invention but to provide guidance to programs and databases which are useful with the nucleic acid codes of the invention or the polypeptide codes of SEQ ID No. 653-654. The programs and databases which may be used include, but are not limited to: MacPattern (EMBL), DiscoveryBase (Molecular Applications Group), GeneMine (Molecular Applications Group), Look (Molecular Applications Group), MacLook (Molecular Applications Group), BLAST and BLAST2 (NCBI), BLASTN and BLASTX (Altschul et al, *J. Mol. Biol.* 215: 403 (1990)), FASTA (Pearson and Lipman, *Proc. Natl. Acad. Sci. USA,* 85: 2444 (1988)), FASTDB (Brutlag et al. Comp. App. Biosci. 6:237-245, 1990), Catalyst (Molecular Simulations Inc.), Catalyst/SHAPE (Molecular Simulations Inc.), Cerius^2.DBAccess (Molecular Simulations Inc.), HypoGen (Molecular Simulations Inc.), Insight II, (Molecular Simulations Inc.), Discover (Molecular Simulations Inc.), CHARMm (Molecular Simulations Inc.), Felix (Molecular Simulations Inc.), DelPhi, (Molecular Simulations Inc.), QuanteMM, (Molecular Simulations Inc.), Homology (Molecular Simulations Inc.), Modeler (Molecular Simulations Inc.), ISIS (Molecular Simulations Inc.), Quanta/Protein Design (Molecular Simulations Inc.), WebLab (Molecular Simulations Inc.), WebLab Diversity Explorer (Molecular Simulations Inc.), Gene Explorer (Molecular Simulations Inc.), SeqFold (Molecular Simulations Inc.), the EMBL/Swissprotein database, the MDL Available Chemicals Directory database, the MDL Drug Data Report data base, the Comprehensive Medicinal Chemistry database, Derwents's World Drug Index database, the BioByteMasterFile database, the Genbank database, and the Genseqn database. Many other programs and

Printed: 29-12-2004 DESC PCT/IB0000184 IB0000184

data bases would be apparent to one of skill in the art given the present disclosure. Motifs which may be detected using the above programs include sequences encoding leucine zippers, helix-turn-helix motifs, glycosylation sites, ubiquitination sites, alpha helices, and beta sheets, signal sequences encoding signal peptides which direct the secretion of the encoded proteins, sequences implicated in transcription regulation such as homeoboxes, acidic stretches, enzymatic active sites, substrate binding sites, and enzymatic cleavage sites.

It should be noted that the nucleic acid codes of the invention further encompass all of the polynucleotides disclosed, described or claimed in the present invention. Also, it should be noted that the polypeptide codes of SEQ ID Nos. 653-654 further encompass all of the polypeptides disclosed, described or claimed in the present invention. Moreover, the present invention specifically contemplates the storage of such codes on computer readable media and computer systems individually or in combination, as well as the use of such codes and combinations in the methods of section "VI. Computer-Related Embodiments."

VII. EXAMPLES

Several of the methods of the present invention are described in the following examples, which are offered by way of illustration and not by way of limitation. Many other modifications and variations of the invention as herein set forth can be made without departing from the spirit and scope thereof and therefore only such limitations should be imposed as are indicated by the appended claims.

Example 1

De Novo Identification Of Biallelic Markers

The biallelic markers set forth in this application were isolated from human genomic sequences. To identify biallelic markers, genomic fragments were amplified, sequenced and compared in a plurality of individuals.

DNA samples

Donors were unrelated and healthy. They represented a sufficient diversity for being representative of a French heterogeneous population. The DNA from 100 individuals was extracted and tested for the *de novo* identification of biallelic markers.

DNA samples were prepared from peripheral venous blood as follows. Thirty ml of peripheral venous blood were taken from each donor in the presence of EDTA. Cells (pellet) were collected after centrifugation for 10 minutes at 2000 rpm. Red cells were lysed in a lysis solution (50 ml final volume: 10 mM Tris pH7.6; 5 mM $MgCl_2$; 10 mM NaCl). The solution was centrifuged (10 minutes, 2000 rpm)

as many times as necessary to eliminate the residual red cells present in the supernatant, after resuspension of the pellet in the lysis solution. The pellet of white cells was lysed overnight at 42°C with 3.7 ml of lysis solution composed of: (a) 3 ml TE 10-2 (Tris-HCl 10 mM, EDTA 2 mM) / NaCl 0.4 M; (b) 200 µl SDS 10%; and (c) 500 µl proteinase K (2 mg proteinase K in TE 10-2 / NaCl 0.4 M).

For the extraction of proteins, 1 ml saturated NaCl (6M) (1/3.5 v/v) was added. After vigorous agitation, the solution was centrifuged for 20 minutes at 10000 rpm. For the precipitation of DNA, 2 to 3 volumes of 100% ethanol were added to the previous supernatant, and the solution was centrifuged for 30 minutes at 2000 rpm. The DNA solution was rinsed three times with 70% ethanol to eliminate salts, and centrifuged for 20 minutes at 2000 rpm. The pellet was dried at 37°C, and resuspended in 1 ml TE 10-1 or 1 ml water. The DNA concentration was evaluated by measuring the OD at 260 nm (1 unit OD = 50 µg/ml DNA). To determine the presence of proteins in the DNA solution, the OD 260 / OD 280 ratio was determined. Only DNA preparations having a OD 260 / OD 280 ratio between 1.8 and 2 were used in the subsequent examples described below. DNA pools were constituted by mixing equivalent quantities of DNA from each individual.

Amplification of genomic DNA by PCR

Amplification of specific genomic sequences was carried out on pooled DNA samples obtained as described above.

Amplification primers

The primers used for the amplification of human genomic DNA fragments were defined with the OSP software (Hillier & Green, 1991). Preferably, primers included, upstream of the specific bases targeted for amplification, a common oligonucleotide tail useful for sequencing. Primers PU contain the following additional PU 5' sequence : TGTAAAACGACGGCCAGT; primers RP contain the following RP 5' sequence : CAGGAAACAGCTATGACC. Primers are listed in Figure 8.

Amplification

PCR assays were performed using the following protocol:

Final volume	25 µl
DNA	2 ng/µl
$MgCl_2$	2 mM
dNTP (each)	200 µM
primer (each)	2.9 ng/µl
Ampli Taq Gold DNA polymerase	0.05 unit/µl
PCR buffer (10x = 0.1 M TrisHCl pH8.3 0.5M KCl)	1x

Printed: 29-12-2004 DESC PCT/IB0000184 IB0000184

DNA amplification was performed on a Genius II thermocycler. After heating at 94°C for 10 min, 40 cycles were performed. Cycling times and temperatures were: 30 sec at 94°C, 55°C for 1 min and 30 sec at 72°C. Holding for 7 min at 72°C allowed final elongation. The quantities of the amplification products obtained were determined on 96-well microtiter plates, using a fluorometer and Picogreen as intercalant agent (Molecular Probes).

Sequencing of amplified genomic DNA and identification of biallelic polymorphisms

Sequencing of the amplified DNA was carried out on ABI 377 sequencers. The sequences of the amplification products were determined using automated dideoxy terminator sequencing reactions with a dye terminator cycle sequencing protocol. The products of the sequencing reactions were run on sequencing gels and the sequences were determined using gel image analysis (ABI Prism DNA Sequencing Analysis software 2.1.2 version).

The sequence data were further evaluated to detect the presence of biallelic markers within the amplified fragments. The polymorphism search was based on the presence of superimposed peaks in the electrophoresis pattern resulting from different bases occurring at the same position. However, the presence of two peaks can be an artifact due to background noise. To exclude such an artifact, the two DNA strands were sequenced and a comparison between the two strands was carried out. In order to be registered as a polymorphic sequence, the polymorphism had to be detected on both strands. Further, some biallelic single nucleotide polymorphisms were confirmed by microsequencing as described below.

Biallelic markers were identified in the analyzed fragments and are shown in Figure 2. Also, the genomic structure of the FLAP gene and 12-LO gene including the relative location of some biallelic markers is shown in Figure 10 and Figure 14, respectively.

Example 2

Genotyping of Biallelic Markers

The biallelic markers identified as described above were further confirmed and their respective frequencies were determined through microsequencing. Microsequencing was carried out on individual DNA samples obtained as described herein.

Microsequencing primers

Amplification of genomic DNA fragments from individual DNA samples was performed as described in Example 1 using the same set of PCR primers. Microsequencing was carried out on the

amplified fragments using specific primers. See Figure 7. The preferred primers used in microsequencing had about 19 nucleotides in length and hybridized just upstream of the considered polymorphic base.

The microsequencing reactions were performed as follows: 5 µl of PCR products were added to 5 µl purification mix (2U SAP (Shrimp alkaline phosphate) (Amersham E70092X)); 2U Exonuclease I (Amersham E70073Z); and 1 µl SAP buffer (200 mM Tris-HCl pH8, 100 mM $MgCl_2$) in a microtiter plate. The reaction mixture was incubated 30 minutes at 37°C, and denatured 10 minutes at 94°C afterwards. To each well was then added 20 µl of microsequencing reaction mixture containing: 10 pmol microsequencing oligonucleotide (19mers, GENSET, crude synthesis, 5 OD), 1 U Thermosequenase (Amersham E79000G), 1.25 µl Thermosequenase buffer (260 mM Tris HCl pH 9.5, 65 mM $MgCl_2$), and the two appropriate fluorescent ddNTPs complementary to the nucleotides at the polymorphic site corresponding to both polymorphic bases (11.25 nM TAMRA-ddTTP ; 16.25 nM ROX-ddCTP ; 1.675 nM REG-ddATP ; 1.25 nM RHO-ddGTP ; Perkin Elmer, Dye Terminator Set 401095). After 4 minutes at 94°C, 20 PCR cycles of 15 sec at 55°C, 5 sec at 72°C, and 10 sec at 94°C were carried out in a Tetrad PTC-225 thermocycler (MJ Research). The microtiter plate was centrifuged 10 sec at 1500 rpm. The unincorporated dye terminators were removed by precipitation with 19 µl $MgCl_2$ 2mM and 55 µl 100 % ethanol. After 15 minute incubation at room temperature, the microtiter plate was centrifuged at 3300 rpm 15 minutes at 4°C. After discarding the supernatants, the microplate was evaporated to dryness under reduced pressure (Speed Vac). Samples were resuspended in 2.5 µl formamide EDTA loading buffer and heated for 2 min at 95°C. 0.8 µl microsequencing reaction were loaded on a 10 % (19:1) polyacrylamide sequencing gel. The data were collected by an ABI PRISM 377 DNA sequencer and processed using the GENESCAN software (Perkin Elmer).

Frequency of biallelic markers

Frequencies are reported for the less common allele only and are shown in Figure 2.

Example 3

Association Study Between Asthma and the Biallelic Markers of the FLAP Gene

Collection of DNA samples from case and control individuals

The disease trait followed in this association study was asthma, a disease involving the leukotriene pathway. The asthmatic population corresponded to 298 individuals that took part in a clinical study for the evaluation of the anti-asthmatic drug Zileuton. More than 90 % of these 298 asthmatic individuals had a Caucasian ethnic background. The control population was composed of

286 individuals from a random US Caucasian population.

Genotyping of case and control individuals

The general strategy to perform the association studies was to individually scan the DNA samples from all individuals in each of the populations described above in order to establish the allele frequencies of the above described biallelic markers in each of these populations.

Allelic frequencies of the above-described biallelic marker alleles in each population were determined by performing microsequencing reactions on amplified fragments obtained by genomic PCR performed on the DNA samples from each individual. Genomic PCR and microsequencing were performed as detailed above in Examples 1 and 2 using the described PCR and microsequencing primers.

Frequency of the biallelic marker alleles of the FLAP gene and association with asthma

Frequencies of biallelic marker alleles were compared in the case-control populations described above. The association curve in Figure 11 shows the p-value obtained for each marker and the localization of the markers in the genomic region harboring the FLAP gene. As shown in Figure 11, the biallelic marker 10-35-390 presented a strong association with asthma, this association being highly significant (pvalue = 2.29×10^{-3}). The two markers 10-32-357 and 10-33-234 showed weak association when tested independently. The biallelic marker 10-35/390 is located in the FLAP gene. Therefore, the association studies results show that a polymorphism of the FLAP gene seems to be related to asthma. The biallelic marker 10-35-390 can be then used in diagnostics with a test based on this marker or on a combination of biallelic markers comprising this marker.

Haplotype frequency analysis

The results of the haplotype analysis using 9 biallelic markers (10-253-298, 10-32-357, 10-33-175, 10-33-234, 10-33-327, 10-35-358, 10-35-390, 12-628-306, and 12-629-241) are shown in Figure 12. Haplotype analysis for association of FLAP markers and asthma was performed by estimating the frequencies of all possible 2, 3 and 4 marker haplotypes in the asthmatic and Caucasian US control populations. Haplotype estimations were performed by applying the Expectation-Maximization (EM) algorithm (Excoffier and Slatkin, 1995), using the EM-HAPLO program (Hawley et al., 1994). Estimated haplotype frequencies in the asthmatic and control populations were compared by means of a chi-square statistical test.

The most significant haplotypes obtained are shown in Figure 12.

Preferred haplotypes comprise either the marker 10-33-234 (allele A) or the marker 10-35-390

(allele T). Preferred haplotype No. 1 (A at 10-33-234 and T at 10-35-390) presented a p-value of $8.2x10^{-4}$ and an odd-ratio of 1.61. Estimated haplotype frequencies were 28.3% in the cases and 19.7 % in the US controls. Also preferred are haplotypes No. 2 (A at 10-33-234 and G at 12-629-241) and haplotype No. 3 (T at 10-33/327 and T at 10-33/390) which presented respectively a p-value of $1.6x10^{-3}$ and $1.8x10^{-3}$, an odd-ratio of 1.65 and 1.53 and haplotypes frequencies of 0.305 and 0.307 for the asthmatic population and of 0.210 and 0.224 for the US control population.

Preferred haplotypes consisting of three markers (haplotype nos. 37, 38, 39 and 41) comprise the marker 10-33-234 (allele A) and the marker 10-35-390 (allele T). Preferred haplotype No. 37 (A at 10-33-234, T at 10-33-390 and C at 12-628-306) presented a p-value of $8.6x10^{-4}$ and an odd-ratio of 1.76. Estimated haplotype frequencies were 26.5 % in the cases and 17.1 % in the US controls. Haplotype No. 40 (A at 10-33-234, C at 12-628-306 and G at 12-629-241) is also very significantly associated with asthma.

Four-marker haplotypes (haplotype Nos. 121 to 125), five-marker haplotypes (haplotype Nos. 247 and 248) and a six-marker haplotype (haplotype No. 373) also showed significant p-values. They all comprise markers 10-33-234 (allele A) and 10-35/390 (allele T), except haplotype no. 124. Other markers in these haplotypes are chosen from the group consisting of 10-235-298 (allele C), 10-35-358 (allele G), 12-628-306 (allele C) and 12-629-241 (allele G).

Haplotype No. 1 is the preferred haplotype of the invention. It can be used in diagnosis of asthma. Moreover, most of the haplotypes significantly associated with asthma comprise the biallelic marker 10-35-390 (allele A) and could also be used in diagnosis.

The statistical significance of the results obtained for the haplotype analysis was evaluated by a phenotypic permutation test reiterated 1000 or 10,000 times on a computer. For this computer simulation, data from the asthmatic and control individuals were pooled and randomly allocated to two groups which contained the same number of individuals as the case-control populations used to produce the data summarized in Figure 12. A haplotype analysis was then run on these artificial groups for the 2 markers included in the haplotype No. 1, which showed the strongest association with asthma. This experiment was reiterated 1000 and 10,000 times and the results are shown in Figure 13. These results demonstrate that among 1000 iterations none and among 10,000 iterations only 1 of the obtained haplotypes had a p-value comparable to the one obtained for the haplotype No. 1. These results clearly validate the statistical significance of the association between this haplotype and asthma.

Example 4

Association Between Asthma And The Biallelic Markers Of The 12-lipoxygenase Gene

Collection of DNA samples from case and control individuals

The disease trait followed in this association study was asthma, a disease involving the leukotriene pathway. The asthmatic population corresponded to 297 individuals that took part in a clinical study for the evaluation of the anti-asthmatic drug zileuton. More than 90 % of these 297 asthmatic individuals had a Caucasian ethnic background. The control population corresponded to 186 individuals from a random US Caucasian population.

Genotyping of case and control individuals

The general strategy to perform the association studies was to individually scan the DNA samples from all individuals in each of the populations described above in order to establish the allele frequencies of the above described biallelic markers in each of these populations.

Allelic frequencies of the above-described biallelic marker alleles in each population were determined by performing microsequencing reactions on amplified fragments obtained by genomic PCR performed on the DNA samples from each individual. Genomic PCR and microsequencing were performed as detailed above in Examples 1 and 2 using the described PCR and microsequencing primers.

Haplotype frequency analysis

None of the single marker alleles showed a significant association with asthma however, significant results were obtained in haplotype studies. Allelic frequencies were useful to check that the markers used in the haplotype studies meet the Hardy-Weinberg proportions (random mating).

Haplotype analysis was performed using 12 biallelic markers and 17 biallelic markers. The results of the haplotype analysis using 12 biallelic markers (12-208-35, 12-226-167, 12-206-366, 10-347-203, 10-347-220, 10-349-97, 10-349-224, 12-196-119, 12-214-129, 12-216-421, 12-219-230 and 12-223-207) are shown in Figure 15. The results of the haplotype analysis using 17 biallelic markers (12-197-244, 12-208-35, 12-226-167, 12-206-366, 10-346-141, 10-347-111, 10-347-165, 10-347-203, 10-347-220, 10-349-97, 10-349-224, 10-341-116, 12-196-119, 12-214-129, 12-216-421 and 12-219-230) are shown in Figure 16. Haplotype analysis for association of 12-LO biallelic markers and asthma was performed by estimating the frequencies of all possible 2, 3 and 4 marker haplotypes in the asthmatic and control populations described above. Haplotype estimations were performed by applying the Expectation-Maximization (EM) algorithm (Excoffier and Slatkin, *Mol. Biol. Evol.*, 12:921-927, 1995),

using the EM-HAPLO program (Hawley et al., *Am. J. Phys. Anthropol.*,18:104, 1994) as described above. Estimated haplotype frequencies in the asthmatic and control population were compared by means of a chi-square statistical test (one degree of freedom).

Figure 15 shows the most significant haplotypes obtained from the 12 biallelic marker analysis. Haplotype No.1 consisting of three biallelic markers (10-347-220, 12-214-129 and 12-219-230) presented a p-value of 2.10^{-5} and an odd-ratio of 3.38. Estimated haplotype frequencies were 12.3% in the cases and 4 % in the controls. Haplotype No.14 consisting of four biallelic markers (10-347-203, 12-196-119, 12-216-421 and 12-219-230) had a p-value of 4.10^{-6} and an odd ratio of 4.18. Estimated haplotype frequencies were 11.8 % in the cases and 3.1% in the controls. Haplotype No.1 and haplotype No.14, are both strongly associated with asthma. Haplotypes Nos. 2-13 and 15-24 also showed very significant Association (see Figure 15).

Figure 16 shows the most significant haplotypes obtained from the 17 biallelic marker analysis. Haplotype No. 1 consisting of two biallelic markers (12-206-366 and 10-349-224) presented a p-value of $1.8\ 10^{-4}$ and an odd-ratio of 2.05. Estimated haplotype frequencies were 42.4 % in the cases and 26.5 % in the controls. Haplotype No. 7 consisting of three biallelic markers (10-349-97, 12-214-129, 12-219-230) had a p-value of $2.3\ 10^{-5}$ and an odd ratio of 3.32. Estimated haplotype frequencies were 12.5 % in the cases and 4.1 % in the controls. Haplotype No. 27 consisting of four biallelic markers (10-349-97, 12-196-119, 12-216-421 and 12-219-230) had a p-value of $5.4\ 10^{-6}$ and an odd ratio of 3.90. Estimated haplotype frequencies were 12.4 % in the cases and 3.5 % in the controls. Haplotypes Nos. 1, 7 and 27 are strongly associated with asthma. Other haplotypes also showed very significant association (see Figure 16).

The statistical significance of the results obtained for the haplotype analysis was evaluated by a phenotypic permutation test reiterated 1000 or 10,000 times on a computer. For this computer simulation, data from the asthmatic and control individuals were pooled and randomly allocated to two groups which contained the same number of individuals as the case-control populations used to produce the data summarized in Figure 15 and Figure 16. A haplotype analysis was then run on these artificial groups for the markers included in haplotype No. 14 from Figure 15 and for the markers included in haplotypes Nos. 7 and 27 from Figure 16, which showed the strongest association with asthma. This experiment was reiterated 1000 and 10,000 times and the results are shown in Figure 19 and Figure 20, respectively. These results demonstrate that among 1000 iterations only 7 and among 10,000 iterations only 39 of the obtained haplotypes from the 12 biallelic marker set had a p-value comparable to the one obtained for haplotype No.14 from Figure 15. Also, among 1000 iterations only 2 of the obtained haplotypes from the 17 biallelic marker set had a p-value comparable to the one obtained for haplotype No. 7 from Figure 16. These results further demonstrate that among 1000 iterations none of the

Printed: 29-12-2004

obtained haplotypes had a p-value comparable to the one obtained for haplotype No. 27 from Figure 16. These results clearly validate the statistical significance of the association between the haplotypes shown in Figures 15 and 16 and asthma.

Example 5

Association between Side Effects upon Treatment with the Anti-Asthmatic Drug Zileuton (Zyflo™) and the Biallelic Markers of the 12-lipoxygenase Gene

Collection of DNA samples from case and control individuals

The side effect examined in this study was the hepatotoxicity experienced by asthmatic individuals as a result of their treatment with Zileuton as part of a clinical study. Asthmatic individuals were unrelated and more than 90% of the individuals had a Caucasian ethnic background. Hepatotoxicity was monitored by measuring the serum levels of alanine aminotransferase (ALT), which is a sensitive indicator of liver cell damage.

More than 90% of the asthmatic individuals participating in this study did not experience Zileuton-associated ALT increase compared to their ALT levels prior to zileuton intake. As mentioned above, an association study is more informative if the case-control populations present extreme phenotypes. Therefore, the asthmatic individuals, which were selected for the side effect positive trait (ALT+), corresponded to 89 individuals that presented at least 3 times the upper limit of normal (ULN) level of ALT. On the other side, the asthmatic individuals that were selected for the side effect negative trait (ALT-) corresponded to 208 individuals that presented less than 1xULN of ALT. ALT+ and ALT- populations corresponded to 4% and 35% respectively of the total asthmatic individuals that participated in this study.

Genotyping of case and control individuals

The general strategy to perform the association studies was to individually scan the DNA samples from all individuals in each of the populations described above in order to establish the allele frequencies of the above described biallelic markers in each of these populations.

Allelic frequencies of the above-described biallelic marker alleles in each population were determined by performing microsequencing reactions on amplified fragments obtained by genomic PCR performed on the DNA samples from each individual. Genomic PCR and microsequencing were performed as detailed above in Examples 1 and 2 using the described PCR and microsequencing primers.

28-02-2000

Haplotype frequency analysis

None of the single marker alleles showed a significant association with hepatoxicity to zileuton, however, significant results were obtained in haplotype studies.

Haplotype analysis was performed using 12 biallelic markers and 17 biallelic markers. The results of the haplotype analysis using 12 biallelic markers (12-208-35, 12-226-167, 12-206-366, 10-347-203, 10-347-220, 10-349-97, 10-349-224, 12-196-119, 12-214-129, 12-216-421, 12-219-230 and 12-223-207) are shown in Figure 17. The results of the haplotype analysis using 17 biallelic markers (12-197-244, 12-208-35, 12-226-167, 12-206-366, 10-346-141, 10-347-111, 10-347-165, 10-347-203, 10-347-220, 10-349-97, 10-349-224, 10-341-116, 12-196-119, 12-214-129, 12-216-421 and 12-219-230) are shown in Figure 18. Haplotype analysis for association of 12-LO biallelic markers and asthma was performed by estimating the frequencies of all possible 2, 3, 4 and 5 marker haplotypes in the ALT+ and ALT- populations described above. Haplotype estimations were performed by applying the Expectation-Maximization (EM) algorithm (Excoffier and Slatkin, *Mol. Biol. Evol.*, 12:921-927, 1995), using the EM-HAPLO program (Hawley et al., *Am. J. Phys. Anthropol.*,18:104, 1994) as described above. Estimated haplotype frequencies in the ALT+ and ALT- populations were compared by means of a chi-square statistical test (one degree of freedom).

Figure 17 shows the most significant haplotypes obtained from the 12 biallelic marker analysis. Haplotype No.3 consisting of three biallelic markers (10-349-224, 12-216-421 and 12-223-207) presented a p-value of 4.10^{-5} and an odd-ratio of 3.53. Estimated haplotype frequencies were 15.1% in the cases and 4.8 % in the controls. Haplotype No. 8 consisting of four biallelic markers (12-206-366, 10-349-224, 12-216-421 and 12-223-207) had a p-value of $2.9.10^{-6}$ and an odd ratio of 4.56. Estimated haplotype frequencies were 15.8 % in the cases and 4% in the controls. Both haplotypes showed strong association with elevated serum ALT level upon treatment with zileuton. Both haplotypes are related as three out of four biallelic marker alleles (T at 10-349-224, A at 12-216-421 and T at 12-223-207) are common to both haplotypes. Haplotypes Nos. 4-7 and 9-25 also showed very significant association.

Figure 18 shows the most significant haplotypes obtained from the 17 biallelic marker analysis. Haplotype No. 11 consisting of three biallelic markers (12-197/244, 10-349-224 and 12-216-421) presented a p-value of $1.7.10^{-3}$ and an odd-ratio of 2.66, for alleles CTA respectively. Estimated haplotype frequencies were 13.7% in the cases and 5.6% in the controls. The p-value obtained by a chi-square distribution with 7 df for this combination of markers is 2.310^{-2} by Omnibus test suggesting that result is highly significant. Another haplotype consisting of four biallelic markers (12-208-35, 10-512/36 ,12-196-119 and 12-219/230) presented a p-value of $3.7.10^{-5}$ and an odd-ratio of 3.74. Estimated haplotype frequencies were 14.7% in the cases and 4.4% in the controls. The p-value obtained by a chi-square distribution with 15 df for this combination of markers is 5.410^{-4} by Omnibus test. Both

haplotypes showed strong association with elevated serum ALT level upon treatment with zileuton. Both haplotypes are related as three out of four biallelic marker alleles (C at 12-197/244, T at 10-349-224 and A at 12-216-421) are common to both haplotypes. Other haplotypes also showed very significant association.

The statistical significance of the results obtained for the haplotype analysis was evaluated by a phenotypic permutation test reiterated 100, 1000 or 10,000 times on a computer. For this computer simulation, data from the ALT+ and ALT- populations were pooled and randomly allocated to two groups which contained the same number of individuals as the ALT+ and ALT- populations used to produce the data summarized in Figure 17 and Figure 18. A haplotype analysis was then run on the artificial groups for the 4 markers included in haplotype No. 8 from Figure 15 and on the artificial groups for the 4 markers included in haplotype No. 13 from Figure 16, which showed the strongest association with secondary effects to zileuton. This experiment was reiterated 1000 and 10,000 times and the results are shown in Figure 19 and Figure 20, resepectively. These results demonstrate that among 1000 iterations only 5 and among 10,000 iterations only 77 of the obtained haplotypes from the 12 biallelic markers had a p-value comparable to the one obtained for haplotype No. 8. These results demonstrate that among 100 iterations only 3 of the obtained haplotypes from the 17 biallelic markers had a p-value comparable to the one obtained for haplotype No. 11. The p-value obtained by permutating affected status for the omnibus LR test is $2.2.10^{-2}$. These results clearly validate the statistical significance of the association between hepatotoxicity to Zyflo™ and the haplotypes Nos. 3-25 and Nos. 6-30 shown in Figure 17 and Figure 18, respectively.

Allele frequency analysis

Allele frequencies were determined in a random US caucasian population, in an asthmatic population showing no side effects upon treatment with Zyflo™ (ALT-) and in an asthmatic population showing elevated alanine aminotransferase levels upon treatment with Zyflo™ (ALT+). Figure 21 is a chart containing a list of preferred 12-LO-related biallelic markers with an indication of the frequency of the least common allele determined by genotyping as described in Example 2.

Example 6

Identification Of Mutations And Of Low Frequency Alleles Of The 12-LO Gene

Exons 6, 8 and 14 of the 12-lipoxygenase gene were screened for mutations by comparing their sequence in individuals exhibiting elevated ALT levels upon treatment with zileuton (ALT+) and in individuals showing normal ALT levels upon treatment with zileuton (ALT-). ALT + and ALT- individuals are further described in Example 5. Intron sequences immediately flanking these exons

were also screened.

To identify mutations, fragments of the 12-LO gene were amplified, sequenced and compared in ALT+ and ALT- individuals. DNA samples from each individual were processed separately.

DNA samples

Individual DNA samples were obtained as described in Example 1.

Amplification of the 12-LO gene

Amplification primers are described in Figure 8. PCR assays were performed as described in Example 1.

Sequencing of amplified genomic DNA: identification of mutations and of low frequency polymorphisms

Sequencing of the amplified DNA was carried out on ABI 377 sequencers. The sequences of the amplification products were determined using automated dideoxy terminator sequencing reactions with a dye terminator cycle sequencing protocol. The products of the sequencing reactions were run on sequencing gels and the sequences were determined using gel image analysis (ABI Prism DNA Sequencing Analysis software 2.1.2 version).

The sequence data was further analyzed to detect the presence of mutations and of low frequency alleles. The sequences of exon 6, exon 8, exon 14 and flanking intronic sequences in 79 ALT+ individuals and 105 ALT- individuals were compared. New polymorphisms/mutations were detected and the genotype of each individual for these markers was determined. Results are shown below:

Marker ID	Position in 12-LO Gene	Least Common Allele/Mutation	Original Allele
10-508-191	5' flanking region	C	T
10-508-245	5' flanking region	T	C
10-511-62	5' flanking region	T	C
10-511-337	5' flanking region	Insertion T	-
10-512-36	5' flanking region	C	G
10-512-318	5' flanking region	A	G
10-513-250	5' flanking region	A	G
10-513-262	5' flanking region	T	C
10-513-352	5' flanking region	A	G
10-513-365	5' flanking region	A	G
10-343-231	Exon 2	Deletion C	-
10-343-366	Intron 2	C	T

10-343-278	Intron 2	T	C
10-343-339	Intron 4	T	G
10-346-23	Intron 4	G	A
10-346-141	Exon 5	A	G
10-346-263	Intron 5	G	C
10-346-305	Intron 5	C	T
10-347-74	Intron 5	A	G
10-347-111	Exon 6	G	C
10-347-165	Exon 6	T	C
10-347-203	Exon 6	G	A
10-347-220	Exon 6	A	G
10-347-271	Intron 6	T	A
10-347-348	Intron 6	A	G
10-348-391	Intron 7	A	G
10-349-47	Intron 7	C	T
10-349-97	Exon 8	G	A
10-349-142	Exon 8	G	C
10-349-216	Exon 8	Deletion CTG	-
10-349-224	Exon 8	T	G
10-349-368	Intron 8	C	T
10-350-72	Intron 8	T	C
10-350-332	Intron 9	C	T
10-507-170	Exon 11	G	A
10-507-321	Intron 11	A	C
10-507-353	Intron 11	T	C
10-507-364	Intron 11	T	C
10-507-405	Intron 11	T	C
10-339-32	Intron 11	T	C
10-339-124	Intron 11	T	C
10-340-112	Exon 13	A	C
10-340-130	Exon 13	A	T
10-340-238	Intron 13	A	G
10-341-116	Exon 14	A	G
10-341-319	Exon 14 (5'UTR)	T	C
10-342-301	3' flanking region	Insertion A	-
10-342-373	3' flanking region	T	C

Low frequency polymorphisms and mutations identified in exons 5, 6, 8, and 13 are associated with amino acid substitutions at the polypeptide level. In each of these amino acid substitutions the original residue is replaced by a non-equivalent amino acid presenting different chemical properties. As a consequence, specificity, activity and function of the 12-LO enzyme are modified. Biallelic marker 10-343-231 is associated with a frame shift in the open reading frame of the 12-LO gene leading to the expression of a variant 12-LO polypeptide comprising only 131 amino acids. This mutant 12-LO enzyme is probably inactive or shows differences in specificity, activity and function. Biallelic marker

10-343-231 is associated with the deletion of a Leu residue in the 12-LO polypeptide.

The mutations and low frequency polymorphisms listed above represent potential functional mutations of the 12-LO gene.

Example 7

Preparation of Antibody Compositions to 12-lipoxygenase Variants

Preferably antibody compositions, specifically binding the 189-His variant of the 12-LO protein or, to the 225-His variant of the 12-LO protein or, to the 243-Cys variant of the 12-LO protein or, to the 261-Arg variant of the 12-LO protein or, to the 322-Asn variant of the 12-LO or, to the 337-Arg variant of the 12-LO protein or to the 574-Lys variant of 12-LO, are prepared. Other preferred antibody compositions of the invention are capable of specifically binding to amino acid positions 110-131 of SEQ ID No. 654.

Substantially pure protein or polypeptide is isolated from transfected or transformed cells containing an expression vector encoding the 12-LO protein or a portion thereof. The concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms per ml. Monoclonal or polyclonal antibodies to the protein can then be prepared as follows:

Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes in the 12-LO protein or a portion thereof can be prepared from murine hybridomas according to the classical method of Kohler and Milstein (*Nature*, 256:495, 1975) or derivative methods thereof (see Harlow and Lane, *Antibodies A Laboratory Manual*, Cold Spring Harbor Laboratory, pp. 53-242, 1988).

Briefly, a mouse is repetitively inoculated with a few micrograms of the 12-LO protein or a portion thereof over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, E., Meth. Enzymol. 70:419 (1980). Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. et al. Basic Methods in Molecular Biology Elsevier, New York. Section 21-2.

Printed: 29-12-2004

Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogeneous epitopes in the 12-LO protein or a portion thereof can be prepared by immunizing suitable non-human animal with the 12-LO protein or a portion thereof, which can be unmodified or modified to enhance immunogenicity. A suitable non-human animal is preferably a non-human mammal is selected, usually a mouse, rat, rabbit, goat, or horse. Alternatively, a crude preparation which, has been enriched for 12-LO concentration can be used to generate antibodies. Such proteins, fragments or preparations are introduced into the non-human mammal in the presence of an appropriate adjuvant (e.g. aluminum hydroxide, RIBI, etc.) which is known in the art. In addition the protein, fragment or preparation can be pretreated with an agent which will increase antigenicity, such agents are known in the art and include, for example, methylated bovine serum albumin (mBSA), bovine serum albumin (BSA), Hepatitis B surface antigen, and keyhole limpet hemocyanin (KLH). Serum from the immunized animal is collected, treated and tested according to known procedures. If the serum contains polyclonal antibodies to undesired epitopes, the polyclonal antibodies can be purified by immunoaffinity chromatography.

Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. Techniques for producing and processing polyclonal antisera are known in the art, see for example, Mayer and Walker (1987). An effective immunization protocol for rabbits can be found in Vaitukaitis, J. et al. *J. Clin. Endocrinol. Metab.* 33:988-991 (1971). Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. et al., Chap. 19 in: Handbook of Experimental Immunology D. Wier (ed) Blackwell (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12 :M). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in: Manual of Clinical Immunology, 2d Ed. (Rose and Friedman, Eds.) Amer. Soc. For Microbiol., Washington, D.C. (1980).

Antibody preparations prepared according to either the monoclonal or the polyclonal protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. The antibodies may also be used in therapeutic compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

28-02-2000

Although this invention has been described in terms of certain preferred embodiments, other embodiments which will be apparent to those of ordinary skill in the art of view of the disclosure herein are also within the scope of this invention. Accordingly, the scope of the invention is intended to be defined only by reference to the appended claims.

In accordance with the regulations relating to Sequence Listings, the following codes have been used in the Sequence Listing to indicate the locations of biallelic markers within the sequences and to identify each of the alleles present at the polymorphic base. The code "r" in the sequences indicates that one allele of the polymorphic base is a guanine, while the other allele is an adenine. The code "y" in the sequences indicates that one allele of the polymorphic base is a thymine, while the other allele is a cytosine. The code "m" in the sequences indicates that one allele of the polymorphic base is an adenine, while the other allele is a cytosine. The code "k" in the sequences indicates that one allele of the polymorphic base is a guanine, while the other allele is a thymine. The code "s" in the sequences indicates that one allele of the polymorphic base is a guanine, while the other allele is a cytosine. The code "w" in the sequences indicates that one allele of the polymorphic base is an adenine, while the other allele is a thymine.

In some instances, the polymorphic bases of the biallelic markers alter the identity of amino acids in the encoded polypeptide. This is indicated in the accompanying Sequence Listing by use of the feature VARIANT, placement of a Xaa at the position of the polymorphic amino acid, and definition of Xaa as the two alternative amino acids. For example, if one allele of a biallelic marker is the codon CAC, which encodes histidine, while the other allele of the biallelic marker is CAA, which encodes glutamine, the Sequence Listing for the encoded polypeptide will contain an Xaa at the location of the polymorphic amino acid. In this instance, Xaa would be defined as being histidine or glutamine.

In other instances, Xaa may indicate an amino acid whose identity is unknown because of nucleotide sequence ambiguity. In this instance, the feature UNSURE is used, Xaa is placed at the position of the unknown amino acid, and Xaa is defined as being any of the 20 amino acids or a limited number of amino acids suggested by the genetic code.

SEQUENCE LISTING FREE TEXT

The following free text appears in the accompanying sequence listing:

arachidonic acid metabolism

Homo Sapiens

allele

polymorphic base

misc_binding

primer_bind

potential

potential complement

potential probe

downstream amplification primer, complement

upstream amplification primer

misc_feature

variant

deletion

insertion

variable motif

5'regulatory region

3'regulatory region

5'UTR

CDS

3'UTR

PRT

exon

WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising a contiguous span of at least 12 nucleotides of a sequence selected from the group consisting of the sequences described in Figure 3 and the complements thereof.

2. A polynucleotide according to claim 1, wherein said span includes an eicosanoid-related biallelic marker in said sequence.

3. An isolated polynucleotide comprising a contiguous span of at least 12 nucleotides of a sequence selected from the group consisting of the sequences described in Figure 4 and the complements thereof, wherein said span includes an eicosanoid-related biallelic marker in said sequence with the alternative allele present at said biallelic marker.

4. An isolated polynucleotide consisting essentially of a contiguous span of 8 to 50 nucleotides of a sequence selected from the group consisting of the sequences described in Figure 4 and the complements thereof, wherein said span includes an eicosanoid-related biallelic marker in said sequence with the original allele present at said biallelic marker.

5. An isolated polynucleotide consisting essentially of a contiguous span of 8 to 50 nucleotides of a sequence selected from the group consisting of the sequences described in Figure 5 and the complements thereof, wherein said span includes an eicosanoid-related biallelic marker in said sequence.

6. A polynucleotide according to any one of claims 2 to 5, wherein said contiguous span is 18 to 35 nucleotides in length and said biallelic marker is within 4 nucleotides of the center of said polynucleotide.

7. A polynucleotide according to claim 6, wherein said polynucleotide consists of said contiguous span and said contiguous span is 25 nucleotides in length and said biallelic marker is at the center of said polynucleotide.

8. A polynucleotide according to claim 1, wherein the 3' end of said contiguous span is present at the 3' end of said polynucleotide.

Printed: 29-12-2004 CLMS IB0000184

9. A polynucleotide according to any one of claims 2 to 5, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide and said biallelic marker is present at the 3' end of said polynucleotide.

10. A polynucleotide according to claim 8, wherein the 3' end of said polynucleotide is located within 20 nucleotides upstream of an eicosanoid-related biallelic marker in said sequence.

11. An isolated polynucleotide consisting essentially of a contiguous span of 8 to 50 nucleotides in a sequence selected from the group consisting of the sequences described in Figure 4, the sequences described in Figure 5, and the complements thereof, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide, and wherein the 3' end of said polynucleotide is located within 20 nucleotides upstream of an eicosanoid-related biallelic marker in said sequence.

12. A polynucleotide according to either claim 10 or 11, wherein the 3' end of said polynucleotide is located 1 nucleotide upstream of an eicosanoid-related biallelic marker in said sequence.

13. A polynucleotide according to claim 1, wherein said polynucleotide consists essentially of a sequence selected from the sequences described in Figure 7.

14. An isolated polynucleotide comprising a contiguous span of at least 12 nucleotides of a sequence selected from the sequences described in Figure 6.

15. A polynucleotide consisting essentially of a sequence selected from the sequences described in Figure 8.

16. A polynucleotide consisting essentially of a sequence selected from the sequences described in Figure 9.

17. A polynucleotide according to any one of claims 1, 3, 4, 5, 11, 14, 15 and 16 wherein said contiguous span comprises at least 15 contiguous nucleotides in said sequence.

18. A polynucleotide according to any one of claims 1, 3, 4, 5, 11, 14, 15 and 16 wherein said contiguous span comprises at least 20 contiguous nucleotides in said sequence.

28-02-2000

19. A polynucleotide according to any one of claims 1, 3, 4, 5, 11, 14, 15 and 16 wherein said contiguous span comprises at least 25 contiguous nucleotides in said sequence.

20. A polynucleotide according to any one of claims 1, 3, 4, 5, 11, 14, 15 and 16 attached to a solid support.

21. An array of polynucleotides comprising at least one polynucleotide according to claim 20.

22. An array according to claim 21, wherein said array is addressable.

23. A polynucleotide according to any one of claims 1, 3, 4, 5, 11, 14, 15 and 16, further comprising a label.

24. A method of genotyping comprising determining the identity of a nucleotide at an eicosanoid-related biallelic marker or 12-LO-related biallelic marker in a biological sample.

25. A method according to claim 24, wherein said eicosanoid-related biallelic marker or 12-LO-related biallelic marker is selected from the biallelic markers described in Figure 2, preferably the biallelic markers found in SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; or more preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652.

26. A method according to claim 24, wherein said biological sample is derived from a single subject.

27. A method according to claim 26, wherein the identity of the nucleotides at said biallelic marker is determined for both copies of said biallelic marker present in said subject's genome.

28. A method according claim 24, wherein said biological sample is derived from multiple subjects.

29. A method according to claim 24, further comprising amplifying a portion of said sequence comprising the biallelic marker prior to said determining step.

30. A method according to claim 29, wherein said amplifying is performed by PCR.

Printed: 29-12-2004 CLMS PCT/IB0000184 IB0000184

31. A method according to claim 24, wherein said determining is performed by a hybridization assay.

32. A method according to claim 24, wherein said determining is performed by a sequencing assay.

33. A method according to claim 24, wherein said determining is performed by a microsequencing assay.

34. A method according to claim 24, wherein said determining is performed by an enzyme-based mismatch detection assay.

35. A method of determining the frequency in a population of an allele of an eicosanoid-related biallelic marker or 12-LO-related biallelic marker, comprising:

a) genotyping individuals from said population for said biallelic marker according to the method of claim 24; and

b) determining the proportional representation of said biallelic marker in said population.

36. A method according to claim 24, wherein said eicosanoid-related biallelic marker or 12-LO-related biallelic marker is selected from the biallelic markers described in Figure 2, preferably the biallelic markers found in SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; or more preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652.

37. A method according to claim 35, wherein said genotyping of step a) is performed on each individual of said population.

38. A method according to claim 35, wherein said genotyping is performed on a single biological sample derived from said population.

39. A method of detecting an association between an allele and a phenotype, comprising the steps of:

28-02-2000

a) determining the frequency of at least one eicosanoid-related biallelic marker allele or 12-LO-related biallelic marker allele in a affected population according to the method of claim 35;

b) determining the frequency of said eicosanoid-related biallelic marker allele or 12-LO-related biallelic marker allele in a control population according to the method of claim 35; and

c) determining whether a statistically significant association exists between said allele and said phenotype.

40. A method of estimating the frequency of a haplotype for a set of biallelic markers in a population, comprising:

a) genotyping each individual in said population for at least one eicosanoid-related biallelic marker or 12-LO-related biallelic marker according to claim 24;

b) genotyping each individual in said population for a second biallelic marker by determining the identity of the nucleotides at said second biallelic marker for both copies of said second biallelic marker present in the genome; and

c) applying a haplotype determination method to the identities of the nucleotides determined in steps a) and b) to obtain an estimate of said frequency.

41. A method according to claim 40, wherein said haplotype determination method is selected from the group consisting of asymmetric PCR amplification, double PCR amplification of specific alleles, the Clark method, or an expectation maximization algorithm.

42. A method according to claim 40, wherein said eicosanoid-related biallelic marker or 12-LO-related biallelic marker is selected from the biallelic markers described in Figure 2, preferably the biallelic markers found in SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; or more preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652.

43. A method of detecting an association between a haplotype and a phenotype, comprising the steps of:

a) estimating the frequency of at least one haplotype in a affected population according to the method of claim 40;

b) estimating the frequency of said haplotype in a control population according to the method of claim 40; and

c) determining whether a statistically significant association exists between said haplotype and said phenotype.

44. A method according to either claim 39 or 43, wherein said control population is a trait negative population.

45 A method according to either claim 39 or 43, wherein said case control population is a random population.

46. A method according to claim 39, wherein each of said genotyping of steps a) and b) is performed on a single pooled biological sample derived from each of said populations.

47. A method according to claim 39, wherein said genotyping of steps a) and b) is performed separately on biological samples derived from each individual in said populations.

48. A method according to either claim 39 or 43, wherein said phenotype is a disease involving arachidonic acid metabolism.

49. A method according to either claim 39 or 43, wherein said phenotype is a response to an agent acting on arachidonic acid metabolism.

50. A method according to either claim 39 or 43, wherein said phenotype is a side effect to an agent acting on arachidonic acid metabolism.

51. A method according to claim 39, wherein the identity of the nucleotides at all of the biallelic markers described in Figure 2 is determined in steps a) and b).

52. A computer readable medium having stored thereon a sequence selected from the group consisting of a nucleic acid code comprising a contiguous span of at least 12 nucleotides of a sequence described in Figure 3, Figure 4, Figure 6 and the complements thereof; wherein said contiguous span of a sequence described in Figure 4 comprises an eicosanoid-related biallelic marker with the alternative allele present at said biallelic marker.

53. A computer system comprising a processor and a data storage device wherein said data storage device has stored thereon a sequence selected from the group consisting of a nucleic acid code comprising a contiguous span of at least 12 nucleotides of a sequence described in Figure 3, Figure 4, Figure 6 and the complements thereof; wherein said contiguous span of a sequence described in Figure 4 comprises an eicosanoid-related biallelic marker with the alternative allele present at said biallelic marker.

54. The computer system of Claim 53 further comprising a sequence comparer and a data storage device having reference sequences stored thereon.

55. The computer system of Claim 54 wherein said sequence comparer comprises a computer program which indicates polymorphisms.

56. A method for comparing a first sequence to a reference sequence, comprising the steps of:

a) reading said first sequence and said reference sequence through use of a computer program which compares sequences; and

b) determining differences between said first sequence and said reference sequence with said computer program; wherein said first sequence is selected from the group consisting of a nucleic acid comprising a contiguous span of at least 12 nucleotides of a sequence described in Figure 3, Figure 4, Figure 6 and the complements thereof; wherein said contiguous span of a sequence described in Figure 4 comprises an eicosanoid-related biallelic marker with the alternative allele present at said biallelic marker.

57. The method of Claim 56, wherein said step b) comprises identifying polymorphisms.

58. A method of administering a drug or treatment comprising:

a) obtaining a nucleic acid sample from an individual;

b) determining the identity of the polymorphic base of at least one eicosanoid-related biallelic marker or 12-LO-related biallelic marker according to the method of claim 29 which is associated with a positive response to said drug or treatment, or at least one eicosanoid-related marker or 12-LO-related biallelic marker or which is associated with a negative response to said drug or treatment; and

c) administering said drug or treatment to said individual if said nucleic acid sample contains at least one biallelic marker associated with a positive response to said drug or

treatment, or if said nucleic acid sample lacks at least one biallelic marker associated with a negative response to said drug or treatment.

59. A method of selecting an individual for inclusion in a clinical trial of a drug or treatment comprising:

a) obtaining a nucleic acid sample from an individual;

b) determining the identity of the polymorphic base of at least one eicosanoid-related biallelic marker or 12-LO-related biallelic marker according to the method of claim 29 which is associated with a positive response to said drug or treatment, or at least one biallelic marker associated with a negative response to said drug or treatment in said nucleic acid sample; and

c) including said individual in said clinical trial if said nucleic acid sample contains at least one biallelic marker which is associated with a positive response to said drug or treatment, or if said nucleic acid sample lacks at least one biallelic marker associated with a negative response to said drug or treatment.

60. A method according to claims 58 or 59, wherein said administering step comprises administering said drug or treatment to said individual if said nucleic acid sample contains at least one biallelic marker associated with a positive response to said drug treatment, and said nucleic acid sample lacks at least one biallelic marker associated with a negative response to said drug or treatment.

61. The method according to either claim 58 or 59, wherein said eicosanoid-related marker or 12-LO-related biallelic marker is selected from the group consisting of the biallelic markers described in Figure 2, preferably the biallelic markers found in SEQ ID Nos. 1-418, 425-489, 491-530, 532-539, and 541-652; or more preferably SEQ ID Nos. 26-418, 425-489, 491-530, 532-539, 541-646, and 651-652.

62. A diagnostic kit comprising a polynucleotide according to any one of claims 2, 3, 4, 5, 10, 11, 13, 14, 15, and 16.

63. The use of a polynucleotide in a hybridization assay for determining the identity of a nucleotide at an eicosanoid-related biallelic marker or 12-LO-related biallelic marker.

64. The use of a polynucleotide in a sequencing assay for determining the identity of a nucleotide at an eicosanoid-related biallelic marker or 12-LO-related biallelic marker.

65. The use of a polynucleotide in an allele specific amplification assay for determining the identity of an eicosanoid-related biallelic marker or 12-LO-related biallelic marker.

66. The use of a polynucleotide in amplifying a segment of nucleotides comprising an eicosanoid-related biallelic marker.

67. An isolated, purified, or recombinant polynucleotide comprising a contiguous span of at least 12 nucleotides of SEQ ID No 651, wherein said contiguous span comprises at least 1 one of the following nucleotide positions of SEQ ID No 651: 1 to 2584, 4425 to 5551, 5634 to 5757, 5881 to 5995, 6100 to 6348, 6510 to 7378, 7523 to 8644, 8855 to 12253, 12341 to12853, 13024 to 13307, 13430 to 16566, 16668 to 16774, 16946 to 17062, 17555 to 20674; and the complements thereof.

68. An isolated, purified, or recombinant polynucleotide comprising a contiguous span of at least 12 nucleotides of SEQ ID No 651 and the complements thereof; wherein said contiguous span comprises at least one nucleotide positions selected from the group consisting of: a C at position 3355, a G at position 3488, a G at position 3489, and a G at position 3708 of SEQ ID No 651.

69. An isolated, purified, or recombinant polynucleotide comprising a contiguous span of at least 12 nucleotides of SEQ ID No 651 and the complements thereof; wherein said contiguous span comprises a least one nucleotide positions selected from the group consisting of: a T at position 2323, a C at position 2341, an A at position 2623, an A at position 2832, a C at position 2844, an A at position 2934, an A at position 2947, a G at position 3802, a G at position 4062, a C at position 4088, a T at position 4109, a T at position 4170, an A at position 6019, a C at position 6375, a C at position 6429, an A at position 6467, a G at position 6484, an A at position 8658, a G at position 8703, an A at position 8777, a G at position 8785, a G at position 13341, an A at position 16836, an A at position 16854, and a T at position 17355 of SEQ ID No 651.

70. An isolated, purified, or recombinant polynucleotide comprising a contiguous span of at least 12 nucleotides of SEQ ID No 652, wherein said contiguous span comprises a T at position 1205 of SEQ ID No 652 or nucleotide positions 2151 to 2157 of SEQ ID No 652; and the complements thereof.

71. An isolated, purified, or recombinant polynucleotide comprising a contiguous span of at least 12 nucleotides of SEQ ID No 652 and the complements thereof; wherein said contiguous span comprises a least one nucleotide position selected from the group consisting of: G at position 366, an A

at position 605, a C at position 712, a T at position 766, an A at position 804, a G at position 821, an A at position 1004, a G at position 1049, an A at position 1123, a G at position 1131, a G at position 1491, an A at position 1742, an A at position 1760, an A at position 1941, and a T at position 2144 of SEQ ID No 652.

72. An isolated, purified, or recombinant polynucleotide consisting of, or consisting essentially of a contiguous span of 8 to 50 nucleotides of either one of SEQ ID Nos: 651, 652 or the complements thereof, wherein said span includes a 12-LO-related biallelic marker in said sequence.

73. A polynucleotide according to claim 72, wherein said 12-LO-related biallelic marker is selected from the group consisting of the biallelic markers described in Table I.

74. A polynucleotide according to claim 72, wherein said 12-LO-related biallelic marker is selected from the group consisting of biallelic markers: 12-197-244, 12-208-35, 12-226-167, 12-206-366, 10-346-141, 10-347-111, 10-347-165, 10-347-203, 10-347-220, 10-349-97, 10-349-224, 10-341-116, 12-196-119, 12-214-129, 12-216-421, 12-219-230, and 12-223-207.

75. An isolated, purified, or recombinant polynucleotide which encodes a polypeptide comprising a contiguous span of at least 6 amino acids of SEQ ID No 653, wherein said contiguous span comprises at least one amino acid position selected from the group consisting of: an His residue et amino acid position 189, an His residue at amino acid position 225, a Cys residue at amino acid position 243, an Arg residue at amino acid position 261, an Asn residue at amino acid position 322, an Arg residue at amino acid position 337, a Asn residue at amino acid position 362, an Asn at amino acid position 568 and a Lys residue at amino acid position 574.

76. An isolated, purified, or recombinant polynucleotide which encodes a polypeptide comprising a contiguous span of at least 6 amino acids of SEQ ID No. 654, wherein said contiguous span comprises at least one of amino acid positions 110-131 of SEQ ID No 654.

77. A recombinant vector comprising a polynucleotide according to any one of claims 67 to 76.

78. A host cell comprising a recombinant vector according to claim 77.

79. A non-human host animal or mammal comprising a recombinant vector according to claim 77.

80. A mammalian host cell comprising a 12-LO gene disrupted by homologous recombination with a knock out vector comprising a polynucleotide according to any one of claims 67 to 76.

81. A non-human host mammal comprising a 12-LO gene disrupted by homologous recombination with a knock out vector comprising a polynucleotide according to any one of claims 67 to 76.

82. A diagnostic kit comprising a polynucleotide according to any one of claims 67 to

83. An isolated, purified, or recombinant polypeptide comprising a contiguous span of at least 6 amino acids of SEQ ID No 653, wherein said contiguous span comprises at least one amino acid position selected from the group consisting of: an His residue et amino acid position 189, an His residue at amino acid position 225, a Cys residue at amino acid position 243, an Arg residue at amino acid position 261, an Asn residue at amino acid position 322, an Arg residue at amino acid position 337, a Asn residue at amino acid position 362, an Asn at amino acid position 568 and a Lys residue at amino acid position 574.

84. An isolated or purified antibody composition capable of selectively binding to an epitope-containing fragment of a polypeptide according to claim 83, wherein said epitope comprises at least one amino acid position selected from the group consisting of: an His residue et amino acid position 189, an His residue at amino acid position 225, a Cys residue at amino acid position 243, an Arg residue at amino acid position 261, an Asn residue at amino acid position 322, an Arg residue at amino acid position 337, a Asn residue at amino acid position 362, an Asn at amino acid position 568 and a Lys residue at amino acid position 574.

85. An isolated, purified, or recombinant polypeptide comprising a contiguous span of at least 6 amino acids of SEQ ID No. 654, wherein said contiguous span comprises at least one of amino acid positions 110-131 of SEQ ID No 654.

86. An isolated or purified antibody composition capable of selectively binding to an epitope-containing fragment of a polypeptide according to claim 85, wherein said epitope comprises at least one of amino acid positions 110-131 of SEQ ID No 654.

Fig.1A

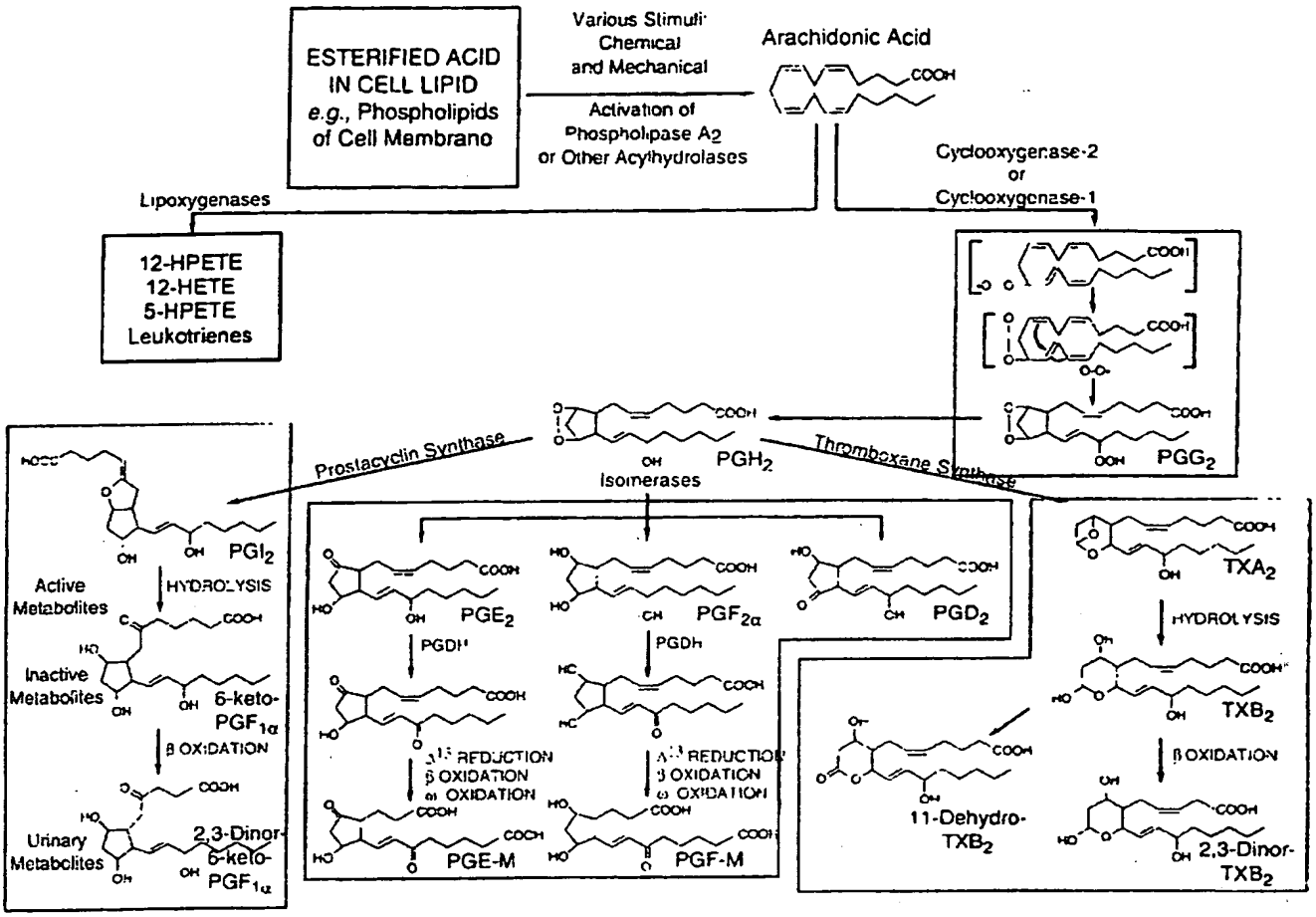

ESTERIFIED ACID IN CELL LIPID e.g., Phospholipids of Cell Membrane
Various Stimuli Chemical and Mechanical
Activation of Phospholipase A2 or Other Acylhydrolases
Arachidonic Acid
Lipoxygenases
Cyclooxygenase-2 or Cyclooxygenase-1
12-HPETE
12-HETE
5-HPETE
Leukotrienes
PGG2
PGH2
Isomerases
Prostacyclin Synthase
Thromboxane Synthase
PGI2
PGE2
PGF2α
PGD2
TXA2
Active Metabolites
HYDROLYSIS
Inactive Metabolites
6-keto-PGF1α
β OXIDATION
Urinary Metabolites
2,3-Dinor-6-keto-PGF1α
PGDH
PGE-M
PGF-M
TXB2
11-Dehydro-TXB2
2,3-Dinor-TXB2

Fig.1B

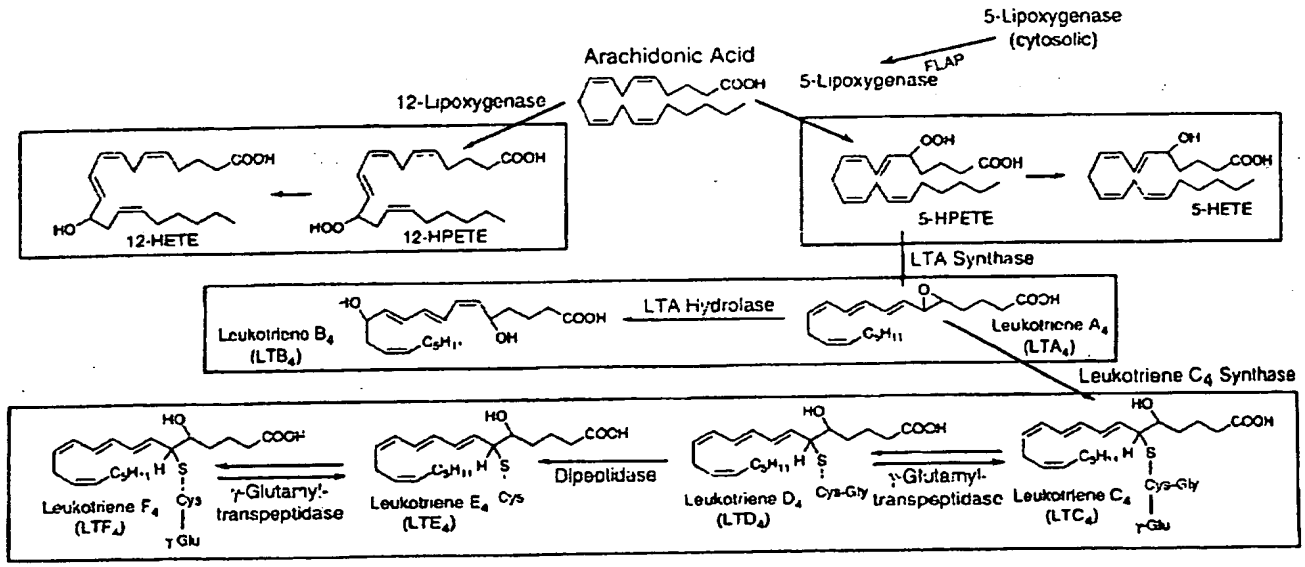

5-Lipoxygenase (cytosolic)
FLAP
5-Lipoxygenase
Arachidonic Acid
12-Lipoxygenase
12-HETE
12-HPETE
5-HPETE
5-HETE
LTA Synthase
LTA Hydrolase
Leukotriene B4 (LTB4)
Leukotriene A4 (LTA4)
Leukotriene C4 Synthase
Leukotriene F4 (LTF4)
γ-Glutamyl-transpeptidase
Leukotriene E4 (LTE4)
Dipeptidase
Leukotriene D4 (LTD4)
γ-Glutamyl-transpeptidase
Leukotriene C4 (LTC4)

FIG.2A

GENE	BIALLELIC MARKER ID	SEQ ID NO.	BIALLELIC MARKER POSITION IN SEQ ID NO.	VALIDATION MICRO-SEQUENCING	GENOTYPING LEAST COMMON ALLELE FREQUENCY %	
FLAP	10-253-118	1	478	N		
FLAP	10-253-298	2	478	Y	G	4.57
FLAP	10-253-315	3	478	N		
FLAP	10-499-155	4	478	N		
FLAP	10-520-256	5	478	N	T	40.8
FLAP	10-500-258	6	478	N		
FLAP	10-500-410	7	478	N		
FLAP	10-503-159	8	478	N		
FLAP	10-504-172	9	478	N		
FLAP	10-504-243	10	478	N		
FLAP	10-204-326	11	478	Y	A	6.63
FLAP	10-32-357	12	478	Y	A	33.5
FLAP	10-33-175	13	478	Y	T	2.30
FLAP	10-33-211	14	478	N		
FLAP	10-33-234	15	478	Y	A	44.0
FLAP	10-33-270	16	478	Y	G/G	
FLAP	10-33-327	17	478	Y	C	24.5
FLAP	10-34-290	18	478	N		
FLAP	10-35-358	19	478	Y	C	31.3
FLAP	10-35-390	20	478	Y	T	23.0
FLAP	10-36-164	21	478	Y	G/G	
FLAP	10-498-192	22	478	N		
FLAP	12-628-306	23	478	Y	T	10.3
FLAP	12-628-311	24	478	N		
FLAP	12-629-241	25	478	Y	C	28.3
12-LO	12-206-366	26	478	Y	C	38.2
12-LO	10-343-339	27	478	N		
12-LO	10-347-74	28	478	N		
12-LO	10-347-111	29	478	N	G/G	
12-LO	10-347-165	30	478	N	C/C	
12-LO	10-347-203	31	478	Y	G	41.6
12-LO	10-347-220	32	478	Y	A	40.5
12-LO	10-347-271	33	478	N		
12-LO	10-347-348	34	478	N		
12-LO	10-348-391	35	478	N		
12-LO	10-349-47	36	478	N		
12-LO	10-349-97	37	478	Y	G	39.6
12-LO	10-349-142	38	478	N	C/C	
12-LO	10-349-224	39	478	Y	T	39.6
12-LO	10-349-368	40	478	N		
12-LO	10-339-32	41	478	N		
12-LO	10-341-116	42	478	Y	A	10.8
12-LO	10-341-319	43	478	N		
12-LO	12-196-119	44	119	Y	C	29.1
12-LO	12-197-244	45	243	Y	C	32.8
12-LO	12-198-128	46	128	N		
12-LO	12-206-81	47	478	N		
12-LO	12-208-35	48	35	Y	A	42.3
12-LO	12-214-129	49	129	Y	C	38.7
12-LO	12-214-151	50	151	N		
12-LO	12-214-360	51	358	N		
12-LO	12-215-467	52	466	N		

FIG.2B

12-LO	12-216-421	53	418	Y	A	36.0
12-LO	12-219-230	54	229	Y	G	32.1
12-LO	12-219-256	55	255	N		
12-LO	12-220-48	56	478	N		
12-LO	12-221-302	57	302	N		
12-LO	12-223-179	58	179	N		
12-LO	12-223-207	59	207	Y	C	38.4
12-LO	12-225-541	60	540	Y	C	37.4
12-LO	12-226-167	61	166	Y	G	41.2
12-LO	12-226-458	62	455	N		
12-LO	12-229-332	63	332	N		
12-LO	12-229-351	64	351	N		
12-LO	12-230-364	65	364	N		
12-LO	12-231-100	66	99	N		
12-LO	12-231-148	67	147	N		
12-LO	12-231-266	68	265	N		
$cPLA_2$	10-231-23	69	500	Y	A	8.79
$cPLA_2$	10-233-386	70	501	Y	G	28.3
$cPLA_2$	10-238-328	71	501	Y	T	17.9
$cPLA_2$	10-239-368	72	501	N		
$cPLA_2$	10-223-30	73	501	Y	G	22.5
$cPLA_2$	10-223-72	74	501	N		
$cPLA_2$	10-223-130	75	501	N		
$cPLA_2$	10-223-262	76	501	N		
$cPLA_2$	10-223-392	77	501	N		
$cPLA_2$	10-224-341	78	501	N		
$cPLA_2$	10-227-282	79	501	Y	G	3.93
ANX1	10-240-241	80	501	N		
ANX1	10-249-185	81	501	N		
ANX1	10-251-128	82	501	N		
ANX1	10-252-209	83	501	N		
ANX1	12-387-32	84	501	Y	G	33.9
ANX1	10-242-316	85	500	N		
ANX1	10-245-412	86	501	N		
ANX1	12-378-171	87	501	N		
ANX1	12-378-228	88	501	N		
ANX1	12-378-450	89	501	N		
ANX1	12-379-65	90	501	N		
ANX1	12-382-204	91	501	Y	G	50.0
ANX1	12-383-117	92	501	N		
ANX1	12-383-170	93	501	N		
ANX1	12-383-268	94	501	N		
ANX1	12-384-336	95	501	N		
ANX1	12-384-451	96	501	N		
ANX1	12-385-123	97	258	N		
ANX1	12-385-427	98	501	N		
ANX1	12-386-155	99	443	Y	G	8.15
ANX1	12-386-24	100	313	N		
ANX1	12-387-177	101	501	Y	T	33.5
ANX1	12-389-431	102	501	N		
ANX1	12-391-366	103	501	N		
ANX1	12-394-85	104	501	N		
ANX1	12-395-382	105	385	N		
ANX1	12-400-217	106	501	Y	G	27.2
ANX1	12-400-280	107	501	N		
ANX1	12-401-378	108	380	N		

FIG.2C

ANX1	12-402-126	109	323	N		
ANX1	12-404-265	110	317	N		
ANX1	12-406-52	111	501	N		
ANX1	12-406-409	112	501	N		
ANX1	12-407-217	113	501	N		
ANX1	12-407-399	114	501	N		
ANX1	12-408-355	115	501	Y	G	2.69
ANX1	12-409-221	116	229	N		
ANX1	12-410-301	117	486	N		
ANX2	10-395-101	118	501	N		
ANX2	10-395-124	119	501	N		
ANX2	10-395-155	120	501	N		
ANX2	10-395-294	121	501	N		
ANX2	10-396-100	122	501	N		
ANX2	10-397-201	123	501	N		
ANX2	10-399-178	124	501	N		
ANX2	10-400-369	125	501	N		
ANX2	10-392-20	126	497	N		
ANX2	10-392-103	127	501	N		
ANX2	10-392-324	128	501	N		
ANX2	10-393-27	129	501	N		
ANX2	10-393-324	130	501	N		
ANX2	12-727-237	131	501	N		
ANX2	12-728-224	132	501	N		
ANX2	12-730-142	133	501	N		
ANX2	12-730-193	134	501	N		
ANX2	12-731-60	135	501	N		
ANX2	12-731-119	136	501	N		
ANX2	12-731-137	137	501	N		
ANX2	12-731-146	138	501	N		
ANX2	12-731-398	139	501	N		
ANX2	12-732-113	140	501	N		
ANX2	12-732-164	141	501	N		
ANX2	12-732-165	142	501	Y	C	27.4
ANX2	12-732-445	143	501	N		
ANX2	12-734-201	144	501	N		
ANX2	12-735-42	145	501	N		
ANX2	12-736-363	146	501	N		
ANX2	12-737-69	147	501	Y	A	36.8
ANX2	12-737-296	148	501	N		
ANX2	12-738-429	149	501	Y	T	35.5
ANX2	12-740-112	150	501	Y	G	37.6
ANX2	12-740-118	151	501	N		
ANX2	12-741-265	152	501	N		
ANX2	12-741-327	153	501	N		
ANX2	12-741-376	154	501	N		
ANX2	12-745-30	155	501	N		
ANX2	12-745-75	156	501	N		
ANX2	12-745-343	157	501	N		
ANX2	12-745-350	158	501	N		
ANX2	12-746-320	159	501	N		
ANX2	12-747-181	160	501	N		
ANX2	12-747-302	161	501	N		
ANX2	12-749-240	162	501	N		
ANX2	12-749-255	163	501	N		
ANX2	12-752-37	164	508	N		

FIG.2D

ANX2	12-752-85	165	501	N		
ANX2	12-752-196	166	501	N		
ANX2	12-752-484	167	501	N		
ANX2	12-753-139	168	501	N		
ANX2	12-753-376	169	501	N		
ANX2	12-754-172	170	501	N		
ANX2	12-754-218	171	501	N		
ANX2	12-754-328	172	501	N		
ANX2	12-754-396	173	501	N		
ANX2	12-755-280	174	501	N		
ANX2	12-757-384	175	501	N		
ANX2	12-758-257	176	501	N		
ANX2	12-758-374	177	501	N		
ANX2	12-758-424	178	501	N		
ANX2	12-761-23	179	541	N		
ANX2	12-761-178	180	501	N		
ANX2	12-764-329	181	501	N		
ANX2	12-764-377	182	501	N		
ANX2	12-765-168	183	501	N		
ANX2	12-765-504	184	501	N		
ANX3	10-372-279	185	501	N		
ANX3	10-375-136	186	501	N		
ANX3	10-376-281	187	501	N		
ANX3	10-369-392	188	501	N		
ANX3	10-371-257	189	501	N		
ANX3	12-513-389	190	501	N		
ANX3	12-513-494	191	501	N		
ANX3	12-515-394	192	501	N		
ANX3	12-516-97	193	501	Y	T	37.2
ANX3	12-520-287	194	501	N		
ANX3	12-520-323	195	501	Y	A	21.5
ANX3	12-523-179	196	501	Y	A	29.9
ANX3	12-523-270	197	501	N		
ANX3	12-527-367	198	501	Y	T	18.9
ANX3	12-529-376	199	501	N		
ANX3	12-529-489	200	501	N		
ANX3	12-530-134	201	501	Y	T	39.3
ANX3	12-530-393	202	501	N		
ANX3	12-531-173	203	501	Y	C	37.6
ANX3	12-539-441	204	501	N		
ANX3	12-543-78	205	501	N		
ANX3	12-543-79	206	501	N		
ANX3	12-546-235	207	501	N		
ANX3	12-549-287	208	501	N		
ANX3	12-550-287	209	501	N		
ANX3	12-552-175	210	501	N		
ANX3	12-554-330	211	501	N		
ANX3	12-556-312	212	501	N		
ANX3	12-556-443	213	501	N		
ANX3	12-558-205	214	501	N		
ANX3	12-558-238	215	501	N		
ANX3	12-558-305	216	501	N		
ANX3	12-769-39	217	501	N		
ANX3	12-769-430	218	501	N		
ANX3	12-770-73	219	501	N		
ANX3	12-772-200	220	501	N		

FIG.2E

ANX3	12-772-254	221	501	N		
CAL1	10-87-73	222	72	N		
CAL1	10-87-74	223	73	N		
CAL1	10-87-80	224	79	N		
CAL1	10-87-140	225	138	N		
CAL1	10-88-81	226	81	Y	C	44.7
CAL1	10-89-41	227	41	N		
CAL1	10-90-35	228	35	Y	A	1.14
CAL1	10-91-274	229	274	N		
CAL1	10-93-133	230	133	N		
CAL1	10-94-197	231	197	Y	G/G	
CAL1	10-94-198	232	198	N		
CAL1	10-166-362	233	362	N		
CAL2	10-207-386	234	387	Y	C/C	
CAL2	10-207-409	235	409	Y	G	9.04
CAL2	10-118-307	236	307	Y	A	0.27
CAL2	10-173-247	237	247	N		
CAL2	10-173-294	238	294	Y	G	2.87
CAL2	10-173-347	239	347	Y	C/C	
CAL3	10-103-104	240	104	N		
CAL3	10-103-323	241	323	Y	T	22.3
CAL3	10-103-402	242	403	N		
CAL3	10-106-98	243	98	N		
CAL3	10-106-288	244	288	Y		
CAL3	10-106-378	245	380	Y		
CAL3	10-168-160	246	160	Y	T	42.1
CAL3	10-168-206	247	206	Y		
CAL3	10-168-284	248	283	N		
CAL3	10-169-318	249	317	N		
CALPA1	12-86-79	250	501	Y	C	37.4
CALPA1	12-88-393	251	501	N		
CALPA1	12-89-369	252	501	Y	G	36.3
CALPA1	12-89-91	253	501	N		
CALPA1	12-94-210	254	501	N		
CALPA1	12-94-516	255	521	N		
CALPA1	12-96-64	256	501	Y	T	8.52
CALPA1	12-97-83	257	501	N		
CALPA1	12-99-296	258	501	Y	T	6.45
CALPA1	12-100-266	259	501	Y	G	32.2
CALPA1	12-811-174	260	501	N		
CALPA1	12-815-94	261	501	N		
CALPA1	12-815-383	262	501	N		
CALPA1	12-815-384	263	500	N		
CALPA1	12-815-391	264	501	N		
CALPA1	12-817-214	265	501	N		
CALPA1	12-817-355	266	501	N		
CALPA1	12-819-437	267	501	N		
CALPA1	12-821-62	268	501	N		
CALPA1	12-821-483	269	501	N		
CALPA1	12-825-173	270	501	N		
CALPA1	12-826-312	271	501	N		
CALPA1	12-831-59	272	501	N		
CALPA1	12-833-264	273	501	N		
CALPA1	12-833-279	274	501	N		
CALPA1	12-833-280	275	502	N		
CALPA1	12-833-373	276	501	N		

FIG.2F

CALPA1	12-834-183	277	483	N		
CALPA1	12-835-54	278	501	N		
CALPA1	12-836-134	279	501	N		
CALPA1	12-836-237	280	500	N		
CALPA1	12-836-238	281	476	N		
CALPA1	12-836-257	282	498	N		
CALPA1	12-836-275	283	501	N		
CALPA1	12-838-179	284	501	N		
CALPA1	12-839-397	285	501	N		
CALPA1	12-840-47	286	501	N		
CALPA1	12-840-77	287	501	N		
CALPA1	12-841-445	288	445	N		
CALPA1	12-842-215	289	501	N		
CALPA1	12-842-447	290	499	N		
CALPA1	12-844-167	291	501	N		
CALPA1	12-845-364	292	501	N		
CALPA1	12-846-209	293	501	N		
CALPA1	12-847-123	294	501	N		
CALPA1	12-849-242	295	501	N		
CYP2J2	10-336-58	296	501	N		
CYP2J2	10-336-137	297	501	N		
CYP2J2	10-336-232	298	501	N		
CYP2J2	12-102-104	299	379	N		
CYP2J2	12-102-111	300	386	N		
CYP2J2	12-102-275	301	501	N		
CYP2J2	12-103-202	302	501	Y	C	14.3
CYP2J2	12-103-214	303	501	N		
CYP2J2	12-104-351	304	501	Y	T	27.4
CYP2J2	12-105-435	305	439	N		
CYP2J2	12-109-149	306	278	Y	A	8.51
CYP2J2	12-109-197	307	326	N		
CYP2J2	12-109-209	308	338	N		
CYP2J2	12-109-284	309	413	N		
CYP2J2	12-113-276	310	501	Y	G	31.2
CYP2J2	12-115-57	311	501	Y	G	8.87
CYP2J2	12-119-26	312	501	Y	G	29.8
COX1	12-347-308	313	501	N		
COX1	12-354-334	314	501	Y	C/C	
COX1	12-357-140	315	501	Y	C	7.14
COX1	12-361-320	316	501	Y	G	18.3
COX1	12-361-388	317	501	Y	A	18.5
COX1	12-365-251	318	501	Y	C	18.8
COX1	12-374-261	319	501	Y	T	21.3
COX1	10-308-116	320	501	N		
COX1	10-311-274	321	501	N		
COX1	10-314-76	322	501	N		
COX1	10-306-265	323	501	N		
COX2	10-52-386	324	386	N		
COX2	10-62-240	325	240	Y	C	12.23
COX2	10-65-276	326	276	Y		
COX2	10-67-42	327	42	N		
COX2	10-67-340	328	341	Y		
COX2	10-55-265	329	264	Y	C	40.9
COX2	10-57-278	330	278	Y		
COX2	10-59-176	331	176	Y		
COX2	10-60-114	332	114	N		

FIG.2G

PGDS	10-27-176	333	176	Y	A	5.32
PGDS	10-28-242	334	242	Y		
PGDS	10-30-349	335	350	Y	A/A	
PGDS	10-181-42	336	42	Y	C	30.2
PGDS	10-181-372	337	374	Y	C	26.3
PGDS	10-183-260	338	259	N		
PG15OH	10-475-163	339	501	N		
PG15OH	12-884-203	340	501	Y	T	29.7
PG15OH	10-479-266	341	501	N		
PG15OH	10-479-350	342	501	N		
PG15OH	10-479-394	343	501	N		
PG15OH	10-482-145	344	501	N		
PG15OH	12-854-64	345	501	N		
PG15OH	12-854-472	346	501	N		
PG15OH	12-855-194	347	501	N		
PG15OH	12-855-288	348	501	N		
PG15OH	12-855-423	349	501	N		
PG15OH	12-857-25	350	476	N		
PG15OH	12-858-346	351	501	Y	T	37.2
PG15OH	12-858-443	352	501	N		
PG15OH	12-860-388	353	501	N		
PG15OH	12-861-270	354	501	N		
PG15OH	12-862-349	355	501	N		
PG15OH	12-862-365	356	501	N		
PG15OH	12-862-452	357	501	N		
PG15OH	12-866-423	358	501	Y	C	46.2
PG15OH	12-867-47	359	501	N		
PG15OH	12-868-181	360	501	N		
PG15OH	12-868-198	361	501	N		
PG15OH	12-868-282	362	501	N		
PG15OH	12-869-128	363	501	N		
PG15OH	12-870-491	364	501	N		
PG15OH	12-872-52	365	501	N		
PG15OH	12-872-293	366	501	N		
PG15OH	12-873-185	367	501	N		
PG15OH	12-873-319	368	501	N		
PG15OH	12-875-248	369	501	Y	G	28.8
PG15OH	12-876-265	370	501	N		
PG15OH	12-876-280	371	501	N		
PG15OH	12-876-454	372	501	N		
PG15OH	12-877-59	373	501	N		
PG15OH	12-877-69	374	501	N		
PG15OH	12-877-79	375	501	N		
PG15OH	12-878-153	376	501	N		
PG15OH	12-878-419	377	501	N		
PG15OH	12-879-67	378	501	N		
PG15OH	12-879-439	379	501	N		
PG15OH	12-881-210	380	501	N		
PG15OH	12-881-389	381	501	N		
PG15OH	12-883-273	382	501	N		
PG15OH	12-885-196	383	501	N		
PG15OH	12-885-333	384	501	N		
PG15OH	12-885-407	385	501	N		
PG15OH	12-885-410	386	501	N		
PG15OH	12-886-195	387	501	Y	A	21.1
PG15OH	12-886-348	388	501	N		

FIG.2H 9/76

PG15OH	12-887-201	389	501	N		
PG15OH	12-887-467	390	501	N		
PG15OH	12-888-98	391	501	N		
PG15OH	12-888-203	392	501	Y	G	38.3
PG15OH	12-888-315	393	501	N		
PG15OH	12-889-518	394	479	N		
PG15OH	12-894-266	395	501	N		
PG15OH	12-895-391	396	501	Y	C	34.6
PG15OH	12-896-140	397	501	N		
PG15OH	12-897-115	398	501	N		
PG15OH	12-897-225	399	501	N		
PG15OH	12-898-49	400	528	N		
CYP8	12-164-119	401	501	Y	C	11.8
CYP8	12-168-84	402	501	Y	T	20.1
CYP8	12-168-365	403	501	N		
CYP8	12-170-299	404	501	Y	T	6.52
CYP8	12-171-360	405	501	Y	T	8.70
CYP8	12-173-59	406	501	Y	G	26.0
CYP8	12-175-214	407	501	Y	A	10.1
CYP8	12-177-183	408	501	Y	G	25.4
CYP8	12-177-366	409	501	N		
TAX2	10-128-45	410	45	Y	T/T	
TAX2	10-128-63	411	63	N		
TAX2	10-123-177	412	177	N		
TAX2	10-123-402	413	402	N		
TAX2	10-120-137	414	136	Y	A	1.60
TAX2	10-120-141	415	140	Y	A	3.09
TAX2	10-179-39	416	39	N		
TAX2	10-180-65	417	65	Y	C	44.7
TAX2	10-179-257	418	257	Y		
15-LOA	10-43-124	419	123	N		
15-LOA	10-43-134	420	133	N		
15-LOA	10-43-193	421	192	N		
15-LOA	10-43-195	422	194	N		
15-LOA	10-43-233	423	232	N		
15-LOA	10-43-138	424	137	Y		
15-LOA	10-46-372	425	369	Y	T	2.43
15-LOA	10-46-36	426	35	N		
15-LOA	10-47-103	427	102	Y		
15-LOA	10-47-125	428	124	Y	T	5.68
15-LOA	10-48-184	429	183	Y	T	28.0
15-LOA	10-48-381	430	382	Y	T	31.4
15-LOA	10-49-33	431	33	Y	T	14.3
15-LOA	10-39-148	432	150	Y	G	14.5
15-LOA	10-40-222	433	222	Y	A	47.6
15-LOA	10-40-252	434	250	N		
15-LOA	10-42-354	435	354	Y		
15-LOA	10-154-42	436	42	N		
15-LOA	10-154-156	437	156	Y	T	24.2
15-LOA	10-154-226	438	226	N		
15-LOB	12-776-259	439	501	N		
5-LO	10-384-109	440	501	N		
5-LO	12-296-388	441	501	Y	G	37.6
5-LO	10-388-379	442	501	N		
5-LO	10-389-116	443	501	N		
5-LO	10-389-349	444	501	N		

FIG.2I

5-LO	10-391-94	445	501	N		
5-LO	12-277-147	446	501	Y	T	44.9
5-LO	12-278-413	447	501	Y	A	33.9
5-LO	12-288-190	448	501	N		
5-LO	12-289-35	449	501	N		
5-LO	12-296-119	450	501	N		
5-LO	12-297-291	451	501	N		
5-LO	12-298-105	452	501	N		
5-LO	12-300-126	453	501	N		
5-LO	12-300-410	454	501	N		
5-LO	12-301-379	455	501	N		
5-LO	12-302-264	456	501	N		
5-LO	12-309-405	457	501	N		
5-LO	12-310-105	458	501	N		
5-LO	12-314-453	459	501	Y	A	18.8
5-LO	12-316-292	460	501	Y	C	40.8
LTA4H	10-281-314	461	501	N		
LTA4H	10-268-381	462	501	N		
LTA4H	12-54-297	463	501	Y	C	9.34
LTA4H	10-276-407	464	501	N		
LTA4H	12-44-50	465	501	Y	A	25.9
LTA4H	12-44-67	466	501	N		
LTA4H	12-45-145	467	501	N		
LTA4H	12-45-166	468	501	N		
LTA4H	12-45-305	469	501	N		
LTA4H	12-46-92	470	501	Y	G	31.9
LTA4H	12-47-132	471	501	Y	C	4.84
LTA4H	12-47-61	472	501	N		
LTA4H	12-48-100	473	501	N		
LTA4H	12-48-323	474	501	N		
LTA4H	12-48-369	475	501	N		
LTA4H	12-48-37	476	501	N		
LTA4H	12-49-131	477	501	Y	A	40.1
LTA4H	12-49-53	478	501	N		
LTA4H	12-49-64	479	501	N		
LTA4H	12-51-234	480	501	Y	A	43.3
LTA4H	12-51-253	481	501	N		
LTA4H	12-51-370	482	501	N		
LTA4H	12-52-400	483	501	N		
LTA4H	12-57-192	484	501	Y	T	41.2
LTA4H	12-57-221	485	501	Y	T	4.40
LTA4H	12-57-510	486	501	N		
LTB4H2	10-1-139	487	139	Y	G	36.3
LTB4H2	10-1-212	488	212	Y	T	16.3
LTB4H2	10-1-241	489	241	Y	A	5.84
LTB4H2	10-9-143	490	143	Y		
LTB4H2	10-9-185	491	185	Y	T/T	
LTB4H2	10-9-264	492	264	Y		
LTB4H2	10-11-22	493	22	N		
LTB4H2	10-13-152	494	152	Y	T	20.8
LTB4H2	10-13-256	495	256	Y		
LTB4H2	10-13-282	496	282	Y	C	25.0
LTB4H2	10-15-281	497	281	N		
LTB4H2	10-17-142	498	142	Y	C/C	
LTB4H2	10-18-302	499	302	N		
LTB4H2	10-23-331	500	331	N		

FIG.2J

LTB4H2	10-25-152	501	152	Y		
LTB4H2	10-25-258	502	258	N		
LTB4H2	10-3-103	503	103	Y	T	47.7
LTB4H2	10-3-144	504	144	Y		
LTB4H2	10-3-275	505	275	Y		
LTB4H2	10-5-227	506	227	Y	A	28.1
LTB4H2	10-7-155	507	155	Y	T	30.4
LTB4H2	10-7-383	508	381	N		
LTB4H2	10-7-98	509	98	N		
LTB412OH	12-561-270	510	501	Y	T	35.2
LTB412OH	12-563-87	511	501	Y	C	28.0
LTB412OH	12-564-64	512	501	Y	T	36.0
LTB412OH	12-564-214	513	501	N		
LTB412OH	12-568-207	514	501	N		
LTB412OH	12-568-365	515	501	N		
LTB412OH	12-568-367	516	501	N		
LTB412OH	12-571-337	517	501	Y	G	17.9
LTB412OH	12-573-378	518	501	Y	A	6.91
LTB412OH	10-294-256	519	501	N		
LTB412OH	10-294-304	520	501	N		
LTB412OH	10-295-201	521	501	N		
LTB412OH	10-296-80	522	501	N		
LTB412OH	10-296-373	523	501	N		
LTB412OH	10-298-122	524	501	N		
LTB412OH	10-298-158	525	501	N		
LTB412OH	10-300-49	526	501	N		
LTB412OH	10-300-185	527	501	N		
LTB4H3	10-10-328	528	327	Y	A	12.5
LTB4H3	10-12-52	529	52	N		
LTB4H3	10-14-46	530	46	Y	T	39.3
LTB4H3	10-19-358	531	357	Y		
LTB4H3	10-20-111	532	110	Y	A	15.8
LTB4H3	10-20-274	533	273	Y	A/A	
LTB4H3	10-24-90	534	90	Y	C	19.2
LTB4H3	10-24-204	535	204	Y	A	25.0
LTB4H3	10-24-221	536	221	N		
LTB4H3	10-24-234	537	234	Y	A	36.1
LTB4H3	10-24-288	538	288	N		
LTB4H3	10-24-311	539	311	N		
LTB4H3	10-26-289	540	289	N		
LTB4H3	10-8-39	541	39	Y		
LTB4H3	10-8-120	542	120	N		
LTB4H3	10-8-154	543	154	N		
LTB4H3	10-8-101	544	101	Y		
LTB4H3	10-8-86	545	86	Y		
LTB4H3	10-8-92	546	92	N		
LTB4H3	10-8-94	547	94	N		
LTB4R	12-61-472	548	501	N		
LTB4R	12-63-402	549	416	N		
LTB4R	12-63-74	550	88	N		
LTB4R	12-64-271	551	287	Y	C	28.6
LTB4R	12-65-98	552	439	N		
LTB4R	12-70-147	553	501	Y	C	11.5
LTB4R	12-70-397	554	501	Y	T	39.7
LTB4R	12-71-320	555	501	Y	A	4.49
LTB4R	12-73-150	556	501	N		

FIG.2K

LTB4R	12-73-49	557	501	Y	A	43.3
LTB4R	12-73-56	558	501	N		
LTB4R	12-74-38	559	501	Y	C	44.1
LTB4R	12-76-238	560	501	Y	T	20.6
LTB4R	12-77-217	561	501	N		
LTB4R	12-77-478	562	501	Y	A	4.40
LTB4R	12-80-114	563	501	N		
LTB4R	12-80-233	564	501	Y	C	4.55
LTB4R	12-82-250	565	250	N		
LTC4	10-176-85	566	85	Y	T	0.54
LTC4	10-176-51	567	51	N		
LTC4	10-176-207	568	207	N		
LTC4	10-176-397	569	397	Y	A	1.63
LTC4	10-177-219	570	219	Y	C	29.0
12-LO	12-214-85	571	85	N		
12-LO	12-215-272	572	271	N		
12-LO	12-221-163	573	163	N		
12-LO	12-225-82	574	82	N		
$cPLA_2$	10-234-179	575	214	Y	Deletion AA	32.6
$cPLA_2$	10-235-272	576	491	N		
ANX1	10-251-342	577	498	N		
ANX2	10-395-367	578	497	N		
ANX2	12-730-58	579	498	N		
ANX2	12-735-208	580	412	Y	Deletion	21.5
ANX2	12-739-22	581	498	Y	Insertion G	23.4
ANX3	12-540-363	582	498	N		
ANX3	12-550-206	583	497	N		
CAL2	12-207-410	584	409	N		
CAL3	10-171-254	585	255	N		
CALPA1	12-94-110	586	498	Y	Deletion AATT	32.5
CALPA1	12-834-290	587	498	N		
COX2	10-55-115	588	114	Y	Deletion TTATA	3.01
PG15OH	12-857-122	589	498	N		
PG15OH	12-872-175	590	498	N		
PG15OH	12-882-40	591	498	N		
PG15OH	12-888-234	592	498	N		
5-LO	12-278-353	593	499	N		
5-LO	12-283-386	594	498	N		
LTA4H	12-44-181	595	458	N		
ANX3	10-370-132	596	501	N		
ANX3	10-370-254	597	501	N		
15PGDHB	10-485-256	598	501	N		
15PGDHB	10-485-257	599	501	N		
15PGDHB	10-474-320	600	501	N		
5LO	10-387-371	601	501	N		
LTB412OH	12-570-239	602	501	N		
LTB412OH	12-570-344	603	501	N		
LTB412OH	12-570-393	604	501	N		
LTB412OH	12-570-421	605	501	N		
LTB412OH	12-570-62	606	501	N		
LTB4H3	10-4-144	607	141	N		
LTB4H3	10-4-161	608	158	N		

FIG.2L

LTB4H3	10-4-270	609	267	N		
LTB4H3	10-4-340	610	337	N		
LTB4H3	10-4-369	611	366	N		
LTB4H3	10-4-420	612	417	N		
LTB4H2	10-13-396	613	396	N		
12-LO	10-509-284	614	501	N		
12-LO	10-509-295	615	501	N		
12-LO	10-339-124	616	501	N		
12-LO	10-340-112	617	501	N		
12-LO	10-340-130	618	501	N		
12-LO	10-340-238	619	501	N		
12-LO	10-342-301	620	501	N		
12-LO	10-342-373	621	501	N		
12-LO	10-343-231	622	501	N		
12-LO	10-343-278	623	501	N		
12-LO	10-346-141	624	501	N	G/G	
12-LO	10-346-23	625	500	N		
12-LO	10-346-263	626	501	N		
12-LO	10-346-305	627	501	N		
12-LO	10-349-216	628	501	N		
12-LO	10-350-332	629	501	N		
12-LO	10-350-72	630	501	N		
12-LO	10-507-170	631	501	N		
12-LO	10-507-321	632	501	N		
12-LO	10-507-353	633	501	N		
12-LO	10-507-364	634	501	N		
12-LO	10-507-405	635	501	N		
12-LO	10-508-191	636	501	N		
12-LO	10-508-245	637	501	N		
12-LO	10-510-173	638	501	N		
12-LO	10-511-337	639	501	N		
12-LO	10-512-36	640	501	Y	C	39.4
12-LO	10-511-62	641	501	N		
12-LO	10-512-318	642	501	N		
12-LO	10-513-250	643	501	N		
12-LO	10-513-262	644	501	N		
12-LO	10-513-352	645	501	N		
12-LO	10-513-365	646	501	N		
FLAP	10-517-100	647	501	N		
FLAP	10-518-125	648	501	N		
FLAP	10-518-194	649	501	N		
FLAP	10-522-71	650	501	N		

FIG.3A

SEQ ID NO.	BIALLELIC MARKER ID	1ST ALLELE	2ND ALLELE	POSITION RANGE OF PREFERRED SEQUENCE
1	10-253-118	A	G	[1-955]
2	10-253-298	G	C	[1-840]
3	10-253-315	C	T	[1-823]
4	10-499-155	A	G	[1-556],[898-955]
5	10-520-256	C	T	[1-384],[726-955]
6	10-500-258	G	T	[1-311],[653-955]
7	10-500-410	A	G	[1-160],[502-955]
8	10-503-159	G	T	[143-160],[388-408],[447-955]
9	10-504-172	A	T	[1-85],[124-792]
10	10-504-243	A	C	[1-15],[54-722]
19	10-35-358	G	C	[555-842]
23	12-628-306	G	A	[1-868],[904-955]
24	12-628-311	T	C	[1-873],[909-955]
25	12-629-241	G	C	[1-17],[247-658],[705-787],[882-955]
27	10-343-339	G	T	[487-506],[733-904]
28	10-347-74	A	G	[1-134],[240-487],[784-956]
35	10-348-391	A	G	[351-552],[682-776]
40	10-349-368	C	T	[416-525]
44	12-196-119	C	T	[1-469]
45	12-197-244	C	T	[153-206]
48	12-208-35	A	T	[1-346],[453-507]
52	12-215-467	G	T	[1-161],[254-499]
53	12-216-421	A	G	[1-486]
54	12-219-230	A	G	[1-485]
55	12-219-256	C	T	[1-485]
56	12-220-48	G	A	[1-577],[883-956]
57	12-221-302	A	C	[1-64],[265-286]
58	12-223-179	A	G	[1-468]
59	12-223-207	C	T	[1-468]
60	12-225-541	C	T	[1-60],[368-598]
61	12-226-167	G	C	[1-255],[344-508]
62	12-226-458	C	T	[1-255],[344-508]
63	12-229-332	G	C	[1-456]
64	12-229-351	G	C	[1-456]
65	12-230-364	C	T	[1-420]
66	12-231-100	C	T	[1-490]
67	12-231-148	C	T	[1-490]
68	12-231-266	C	T	[1-490]
71	10-238-328	C	T	[1-254],[483-728]
72	10-239-368	C	T	[1-144],[373-618]
73	10-223-30	G	C	[1-653],[729-1001]
74	10-223-72	A	G	[1-612],[688-1001]
75	10-223-130	A	T	[1-555],[631-1001]
76	10-223-262	A	G	[1-424],[500-1001]
77	10-223-392	A	G	[1-294],[370-1001]
78	10-224-341	C	T	[137-176],[428-563],[920-1001]
82	10-251-128	A	G	[202-240],[373-415],[464-518],[581-777]
84	12-387-32	A	G	[1-396],[464-1001]
85	10-242-316	G	C	[1-350],[418-1000]
86	10-245-412	A	G	[367-701]
87	12-378-171	T	C	[1-731]
88	12-378-228	G	A	[1-788]
89	12-378-450	T	A	[1-1001]
90	12-379-65	A	G	[1-1001]
91	12-382-204	A	G	[1-1001]

FIG.3B

92	12-383-117	A	G	[1-37],[246-317],[383-1001]
93	12-383-170	A	G	[193-264],[330-1001]
94	12-383-268	G	T	[95-166],[232-1001]
98	12-385-427	G	T	[257-826]
99	12-386-155	G	T	[272-682],[823-943]
100	12-386-24	C	T	[272-682]
101	12-387-177	C	T	[1-251],[319-1001]
102	12-389-431	C	T	[1-386],[470-583],[644-996]
103	12-391-366	C	T	[293-1001]
104	12-394-85	A	C	[1-103],[184-266],[345-1001]
105	12-395-382	A	G	[1-885]
108	12-401-378	A	G	[1-880]
109	12-402-126	C	T	[99-823]
110	12-404-265	A	G	[1-261],[314-501],[715-733],[782-817]
111	12-406-52	C	T	136-952],[984-1001]
112	12-406-409	A	G	[1-595],[627-1001]
113	12-407-217	G	C	[247-673]
114	12-407-399	A	T	[1-491],[955-1001]
115	12-408-355	G	C	[80-907]
116	12-409-221	A	C	[1-500]
117	12-410-301	C	T	[111-986]
118	10-395-101	A	G	[1-529],[633-1001]
119	10-395-124	A	G	[1-539],[611-1001]
120	10-395-155	A	T	[1-509],[581-1001]
121	10-395-294	C	T	[1-371],[443-858]
122	10-396-100	A	G	[1-506],[635-776],[952-1001]
124	10-399-178	A	G	[1-142],[178-514],[632-1001]
125	10-400-369	A	T	[1-285],[385-513],[555-844],[878-941]
126	10-392-20	A	G	[75-203],[245-534],[568-631],[746-849],[898-997]
127	10-392-103	A	G	[1-552],[663-770],[819-1001]
128	10-392-324	G	C	[1-331],[442-549],[598-891],[977-1001]
129	10-393-27	G	C	[1-76],[187-294],[343-636],[722-1001]
130	10-393-324	A	G	[1-340],[423-1001]
131	12-727-237	A	G	[513-1001]
132	12-728-224	A	G	[352-507],[661-772],[862-1001]
133	12-730-142	A	G	[1-1001]
134	12-730-193	A	G	[1-1001]
135	12-731-60	C	T	[97-665],[711-729],[898-1001]
136	12-731-119	C	T	[1-606],[652-670],[839-1001]
137	12-731-137	G	T	[1-588],[634-652],[821-1001]
138	12-731-146	A	C	[1-579],[625-643],[812-1001]
139	12-731-398	C	T	[1-327],[373-391],[560-743],[823-1001]
140	12-732-113	A	G	[58-1001]
141	12-732-164	A	G	[1-1001]
142	12-732-165	G	C	[1-1001]
143	12-732-445	C	T	[1-935],[975-1001]
144	12-734-201	T	C	[161-1001]
145	12-735-42	G	A	[1-343],[374-566],[656-682],[731-961]
146	12-736-363	G	A	[1-1001]
147	12-737-69	T	C	[1-739]
148	12-737-296	G	A	[1-960]
149	12-738-429	G	A	[1-205],[411-1001]
150	12-740-112	A	G	[1-26],[144-616],[743-1001]
151	12-740-118	C	T	[1-20],[138-610],[737-1001]
152	12-741-265	G	A	[1-1001]
153	12-741-327	T	A	[1-1001]
154	12-741-376	G	A	[1-1001]

FIG.3C

155	12-745-30	G	A	[1-1001]
156	12-745-75	T	C	[1-1001]
157	12-745-343	T	G	[1-1001]
158	12-745-350	C	A	[1-1001]
159	12-746-320	C	T	[1-1001]
160	12-747-181	C	T	[1-1001]
161	12-747-302	C	T	[1-1001]
164	12-752-37	G	A	[1-1003]
165	12-752-85	C	G	[1-1001]
166	12-752-196	T	C	[1-62],[108-1001]
167	12-752-484	T	C	[396-1001]
168	12-753-139	C	T	[1-1001]
169	12-753-376	C	T	[1-778],[855-1001]
170	12-754-172	C	T	[1-1001]
171	12-754-218	C	T	[1-1001]
172	12-754-328	G	C	[1-1001]
173	12-754-396	G	T	[1-1001]
174	12-755-280	G	C	[1-1001]
176	12-758-257	A	C	[1-1001]
177	12-758-374	A	C	[1-1001]
178	12-758-424	A	G	[1-1001]
179	12-761-23	G	A	[1-177],[253-701]
180	12-761-178	G	A	[1-292],[368-1001]
181	12-764-329	G	A	[1-1001]
182	12-764-377	G	A	[1-1001]
183	12-765-168	G	A	[1-906]
184	12-765-504	T	C	[1-1002]
190	12-513-389	C	T	[1-1001]
191	12-513-494	G	C	[1-999]
192	12-515-394	A	T	[77-950]
193	12-516-97	C	T	[1-744],[798-1001]
194	12-520-287	A	T	[179-468],[506-885]
195	12-520-323	A	G	[143-432],[470-849]
196	12-523-179	G	A	[1-291],[344-1001]
197	12-523-270	G	A	[1-382],[435-1001]
198	12-527-367	T	A	[1-496],[595-1001]
199	12-529-376	T	C	[279-1001]
200	12-529-489	T	C	[1-37],[391-1001]
201	12-530-134	A	T	[1-94],[166-224],[316-803]
202	12-530-393	C	T	[57-544],[766-1001]
203	12-531-173	C	T	[1-231],[414-735],[789-1001]
204	12-539-441	C	T	[1-1001]
205	12-543-78	G	A	[1-836]
206	12-543-79	C	G	[1-837]
207	12-546-235	C	T	[1-403],[492-1001]
208	12-549-287	T	C	[149-494]
209	12-550-287	A	G	[304-1001]
210	12-552-175	G	A	[1-750],[831-883]
211	12-554-330	G	T	[1-1001]
212	12-556-312	A	C	[1-1001]
213	12-556-443	C	T	[1-1001]
214	12-558-205	C	G	[1-1001]
215	12-558-238	T	C	[1-1001]
216	12-558-305	T	A	[1-1001]
217	12-769-39	G	T	[1-292],[593-624],[690-1001]
218	12-769-430	C	T	[202-233],[299-633]
219	12-770-73	G	A	[1-716]

FIG.3D

220	12-772-200	G	A	[1-732],[788-1001]
221	12-772-254	T	C	[1-786],[842-1001]
233	10-166-362	A	C	
250	12-86-79	G	A	[70-653],[748-1001]
252	12-89-369	G	C	[1-51],[102-1001]
253	12-89-91	A	G	[1-329],[380-1000]
254	12-94-210	C	T	[573-588]
255	12-94-516	A	T	[287-302]
256	12-96-64	C	A	[1-630],[936-1001]
257	12-97-83	A	C	[1-20],[543-649],[719-916],[964-1001]
258	12-99-296	G	A	[1-210],[305-522],[904-1001]
259	12-100-266	T	C	[504-545],[927-949]
260	12-811-174	T	C	[1-945]
261	12-815-94	A	G	[1-1001]
262	12-815-383	A	G	[1-1001]
263	12-815-384	G	C	[1-1001]
264	12-815-391	C	T	[1-1001]
268	12-821-62	T	G	[1-294],[376-437],[621-887]
269	12-821-483	T	G	[1-48],[460-510],[664-715],[797-858]
270	12-825-173	A	C	[1-34],[522-1001]
273	12-833-264	T	A	[1-86],[216-446],[558-1001]
274	12-833-279	G	A	[1-101],[231-461],[573-1001]
275	12-833-280	T	C	[1-102],[232-462],[574-1001]
276	12-833-373	G	A	[1-195],[325-555],[667-1001]
277	12-834-183	A	G	[295-990]
278	12-835-54	A	G	[1-1001]
279	12-836-134	C	T	[84-249],[354-587],[633-1001]
280	12-836-237	A	G	[1-147],[252-945]
281	12-836-238	A	T	[1-123],[228-919]
282	12-836-257	A	G	[1-123],[228-919]
283	12-836-275	A	C	[1-108],[213-904]
284	12-838-179	A	G	[1-519],[718-1001]
285	12-839-397	G	A	[1-43],[110-1001]
286	12-840-47	C	G	[1-553],[659-1001]
287	12-840-77	T	C	[1-583],[689-1001]
288	12-841-445	G	C	[1-502]
291	12-844-167	T	C	[186-1001]
292	12-845-364	G	A	[1-849]
293	12-846-209	A	T	[1-817]
294	12-847-123	A	G	[1-1001]
295	12-849-242	C	A	[1-27],[490-658]
298	10-336-232	A	G	[507-1001]
299	12-102-104	A	G	[1-630],[712-790]
300	12-102-111	A	G	[1-630],[712-790]
301	12-102-275	A	G	[1-581],[663-741],[834-851],[891-1001]
302	12-103-202	C	T	[188-767]
303	12-103-214	A	G	[176-755]
304	12-104-351	T	G	[1-20],[336-402],[438-511],[911-935]
305	12-105-435	A	G	[1-147],[492-924]
306	12-109-149	A	G	[1-59],[289-607]
307	12-109-197	A	G	[1-59],[289-607]
308	12-109-209	A	G	[1-59],[289-607]
309	12-109-284	A	G	[1-59],[289-607]
310	12-113-276	T	C	[1-1001]
311	12-115-57	A	G	[507-1001]
312	12-119-26	T	C	[1-569]
314	12-354-334	G	A	[1-750]

FIG.3E

315	12-357-140	C	T	[1-1001]
316	12-361-320	G	T	[1-201],[268-1001]
317	12-361-388	A	G	[1-133],[200-1001]
318	12-365-251	G	C	[1-41],[132-151],[232-622],[688-933]
319	12-374-261	G	A	[249-1001]
321	10-311-274	C	T	[125-305],[472-878]
322	10-314-76	C	T	[1-224],[290-535],[803-1001]
335	10-30-349	A	G	
340	12-884-203	C	T	[1-349],[464-1001]
342	10-479-350	C	T	[1-280],[446-1001]
343	10-479-394	A	G	[1-236],[402-1001]
345	12-854-64	A	G	[1-1001]
346	12-854-472	G	T	[1-1001]
347	12-855-194	T	G	[1-1001]
348	12-855-288	T	C	[1-1001]
349	12-855-423	T	G	[1-1001]
350	12-857-25	C	T	[221-985]
351	12-858-346	T	C	[1-1001]
352	12-858-443	G	A	[1-1001]
353	12-860-388	G	A	[1-30],[157-628],[831-1001]
354	12-861-270	C	T	[1-780]
355	12-862-349	A	G	[78-1001]
356	12-862-365	C	T	[62-1001]
357	12-862-452	G	T	[1-1000]
358	12-866-423	C	T	[1-434],[521-1001]
359	12-867-47	C	T	[81-769]
360	12-868-181	A	G	[306-1001]
361	12-868-198	A	G	[289-1001]
362	12-868-282	C	T	[205-1001]
363	12-869-128	A	C	[1-128],[908-1001]
365	12-872-52	A	G	[436-1001]
366	12-872-293	A	G	[185-1001]
367	12-873-185	T	C	[114-257],[288-377],[572-1001]
368	12-873-319	T	A	[1-139],[248-391],[422-511],[706-1001]
369	12-875-248	T	C	[1-408],[525-1001]
370	12-876-265	T	A	[1-1001]
371	12-876-280	C	G	[1-1001]
372	12-876-454	G	A	[1-1001]
373	12-877-59	C	T	[329-1001]
374	12-877-69	G	T	[319-1001]
375	12-877-79	C	T	[309-1001]
376	12-878-153	C	T	[207-937]
377	12-878-419	G	T	[1-629],[734-929]
378	12-879-67	G	C	[1-200],[261-460],[527-1001]
379	12-879-439	A	G	[1-89],[156-796]
380	12-881-210	A	G	[1-1001]
381	12-881-389	G	T	[1-841]
382	12-883-273	G	C	[1-56],[96-1001]
383	12-885-196	T	C	[1-1001]
384	12-885-333	C	G	[1-1001]
385	12-885-407	T	C	[1-1001]
386	12-885-410	C	G	[1-1001]
387	12-886-195	T	C	[1-815],[867-1001]
388	12-886-348	T	C	[1-968]
389	12-887-201	G	A	[1-59],[181-1001]
390	12-887-467	T	C	[295-325],[447-1001]
391	12-888-98	G	A	[1-717],[916-1001]

FIG.3F

392	12-888-203	C	A	[1-822]
393	12-888-315	T	G	[1-1001]
394	12-889-518	G	A	[1-89],[280-320],[441-1001]
395	12-894-266	T	C	[1-1001]
396	12-895-391	G	A	[148-1001]
397	12-896-140	T	A	[60-76],[126-1001]
398	12-897-115	T	C	[259-557]
399	12-897-225	G	A	[369-667]
400	12-898-49	G	A	[1-283],[372-781]
401	12-164-119	T	G	[1-646],[979-1001]
403	12-168-365	C	G	[1-600]
407	12-175-214	A	G	[1-154],[227-317],[391-660],[747-1001]
408	12-177-183	C	G	[1-837],[975-1001]
409	12-177-366	C	A	[1-1001]
427	10-47-103	A	C	
428	10-47-125	A	T	
433	10-40-222	A	G	
434	10-40-252	C	T	
442	10-388-379	C	T	[1-202],[383-1001]
443	10-389-116	A	G	[1-538],[693-1001]
444	10-389-349	C	T	[1-305],[460-1001]
445	10-391-94	A	G	[1-259],[301-575],[691-928]
446	12-277-147	A	T	[1-693]
447	12-278-413	A	G	[1-151],[365-733],[775-1001]
448	12-288-190	G	A	[1-701]
449	12-289-35	A	G	[1-791],[946-1001]
450	12-296-119	A	G	[451-550]
451	12-297-291	C	T	[1-1001]
452	12-298-105	G	A	[1-162],[348-1001]
453	12-300-126	A	G	[1-782]
454	12-300-410	A	C	[1-415],[447-498]
455	12-301-379	A	T	[1-627],[932-1001]
456	12-302-264	G	A	[1-1001]
458	12-310-105	G	C	[293-1001]
459	12-314-453	A	T	[1-392],[439-558],[643-799]
460	12-316-292	C	T	[1-460]
461	10-281-314	G	T	[1-282],[453-832],[921-1001]
462	10-268-381	C	T	[1-197],[383-895]
463	12-54-297	C	T	[97-326],[404-518],[658-1001]
464	10-276-407	C	T	[1-97],[510-615],[954-1001]
465	12-44-50	T	C	[220-534],[918-1001]
466	12-44-67	T	C	[237-551],[935-1001]
469	12-45-305	C	T	[1-63],[488-816]
470	12-46-92	A	G	[83-1001]
471	12-47-132	C	T	[1-184],[457-685],[799-871],[987-1001]
472	12-47-61	C	T	[72-255],[528-756],[870-942]
473	12-48-100	A	G	[1-1001]
474	12-48-323	A	G	[1-747]
475	12-48-369	C	T	[1-682]
476	12-48-37	C	T	[1-1001]
477	12-49-131	T	C	[1-609],[677-749],[920-1001]
478	12-49-53	G	A	[1-531],[599-671],[842-1001]
479	12-49-64	G	A	[1-542],[610-682],[853-1001]
480	12-51-234	T	C	[1-47],[182-541],[919-1001]
481	12-51-253	C	A	[1-66],[201-560],[938-1001]
482	12-51-370	G	A	[1-182],[317-676]
483	12-52-400	G	A	[1-100],[404-1001]

FIG.3G

484	12-57-192	G	A	[1-168],[286-752]
485	12-57-221	G	A	[1-197],[315-781]
486	12-57-510	C	A	[1-163],[251-486],[604-1010]
494	10-13-152	C	T	
510	12-561-270	C	T	[188-203],[496-642],[697-738]
511	12-563-87	C	T	[1-929]
512	12-564-64	G	T	[1-213],[381-1001]
513	12-564-214	C	T	[1-64],[232-1001]
514	12-568-207	G	T	[424-513],[613-1001]
515	12-568-365	G	T	[266-355],[455-1001]
516	12-568-367	G	T	[264-353],[453-1001]
517	12-571-337	G	C	[1-53],[327-897]
518	12-573-378	A	G	[1-335],[437-910]
519	10-294-256	G	C	[1-53],[327-897]
520	10-294-304	G	C	[279-849],[942-1001]
522	10-296-80	A	G	[359-397],[531-906]
523	10-296-373	A	G	[60-105],[239-623],[924-1001]
524	10-298-122	C	T	[1-565],[737-873]
525	10-298-158	A	G	[1-529],[701-837]
526	10-300-49	A	G	[285-643],[808-854]
527	10-300-185	C	T	[92-507],[672-718],[976-1001]
549	12-63-402	A	G	[1-472]
550	12-63-74	A	G	[1-472]
551	12-64-271	C	T	[1-787]
552	12-65-98	C	T	[112-272],[334-864]
553	12-70-147	A	C	[1-211],[491-1001]
554	12-70-397	C	T	[241-1001]
555	12-71-320	A	G	[1-1001]
556	12-73-150	C	T	[1-140],[275-607],[646-821]
557	12-73-49	A	G	[1-240],[375-707],[746-921]
558	12-73-56	A	T	[1-233],[368-700],[739-914]
559	12-74-38	G	A	[1-1001]
561	12-77-217	C	T	[1-822]
562	12-77-478	A	G	[1-562]
563	12-80-114	T	C	[1-1001]
564	12-80-233	G	A	[1-1001]
565	12-82-250	A	T	[404-454]
571	12-214-85	CCTAT	-	[1-101],[259-305]
572	12-215-272	T	-	[1-161],[254-499]
573	12-221-163	GTCCTCA	T	[1-64],[265-286]
574	12-225-82	T	-	[1-60],[368-598]
577	10-251-342	GG	C	[1-56],[156-301],[364-560]
578	10-395-367	A	-	[1-263],[367-717],[764-783]
579	12-730-58	ACAA	-	[162-251],[287-321],[517-767]
580	12-735-208	-	Deletion	[1-689],[779-805],[854-1002]
581	12-739-22	G	-	[1-39],[386-640],[791-1002]
582	12-540-363	T	-	[1-1002]
583	12-550-206	T	-	[380-1002]
587	12-834-290	G	-	[196-1002]
589	12-857-122	CTCT	-	[145-1002]
590	12-872-175	T	-	[1-41],[310-1102]
592	12-888-234	C	-	[1-850],[950-1002]
593	12-278-353	A	-	[1-208],[422-790],[832-1001]
595	12-44-181	C	-	[308-622],[983-1002]
602	12-570-239	T	C	[386,671],[724,727],[947,1001]
603	12-570-344	T	C	[1,51],[491,601],[727,776],[829,832]

FIG.3H

619	10-340-238	A	G	[231,310],[487,601]
620	10-342-301	Insertion	-	[432,576],[605,609],[676,722]
621	10-342-373	C	T	[360,504],[533,537],[604,650],[930,1001]
625	10-346-23	A	G	[1,144],[233,274],[305,347],[478,592],[696,945]
626	10-346-263	G	C	[1,37],[68,110],[241,355],[459,708]
627	10-346-305	C	T	[1,68],[199,313],[417,666],[961,1001]
629	10-350-332	C	T	[1,913]
630	10-350-72	C	T	[1,1001]
632	10-507-321	A	C	[1,308],[440,462],[552,652],[711,1000]
633	10-507-353	C	T	[1,276],[408,430],[520,620],[679,1000]
634	10-507-364	C	T	[1,265],[397,609],[668,1000]
635	10-507-405	C	T	[1,224],[356,378],[468,568],[627,1000]
636	10-508-191	C	T	[1,403],[442,444],[491,640],[942,1000]
637	10-508-245	C	T	[1,349],[388,390],[463,586],[888,1000]
638	10-510-173	ATTTA	TTTTTT	[243,380],[411,546]
647	10-517-100	G	C	[1,1000]
648	10-518-125	G	T	[1,1000]
649	10-518-194	A	G	[1,1000]
650	10-522-71	A	G	[1,806],[844,863],[911,920],[950,1000]

FIG.4A

SEQ ID NO.	BIALLELIC MARKER ID	ORIGINAL ALLELE	ALTERNATIVE ALLELE
11	10-204-326	G	A
12	10-32-357	C	A
13	10-33-175	C	T
14	10-33-211	C	T
15	10-33-234	A	C
16	10-33-270	G	A
17	10-33-327	T	C
18	10-34-290	G	T
20	10-35-390	C	T
21	10-36-164	G	A
26	12-206-366	T	C
29	10-347-111	G	C
30	10-347-165	C	T
33	10-347-271	A	T
34	10-347-348	G	A
36	10-349-47	T	C
38	10-349-142	C	G
41	10-339-32	C	T
43	10-341-319	C	T
46	12-198-128	G	A
47	12-206-81	G	A
49	12-214-129	C	T
50	12-214-151	G	C
51	12-214-360	G	C
69	10-231-23	G	A
70	10-233-386	A	G
79	10-227-282	A	G
80	10-240-241	A	G
83	10-252-209	G	A
95	12-384-336	C	T
96	12-384-451	G	C
97	12-385-123	C	T
106	12-400-217	A	G
107	12-400-280	A	G
162	12-749-240	G	A
163	12-749-255	G	T
175	12-757-384	T	C
185	10-372-279	T	C
186	10-375-136	T	C
187	10-376-281	A	T
188	10-369-392	C	T
222	10-87-73	C	T
223	10-87-74	A	T
224	10-87-80	A	G
225	10-87-140	C	T
226	10-88-81	T	C
227	10-89-41	G	A
228	10-90-35	G	A
229	10-91-274	T	G
231	10-94-197	G	A
232	10-94-198	T	G
234	10-207-386	C	G
235	10-207-409	G	C
236	10-118-307	G	A
237	10-173-247	G	A

FIG.4B

238	10-173-294	A	G
239	10-173-347	C	T
240	10-103-104	C	T
241	10-103-323	T	C
242	10-103-402	C	T
243	10-106-98	C	A
246	10-168-160	T	C
247	10-168-206	C	A
248	10-168-284	T	A
249	10-169-318	C	A
251	12-88-393	A	C
265	12-817-214	G	A
266	12-817-355	T	C
267	12-819-437	A	G
271	12-826-312	G	A
272	12-831-59	G	C
289	12-842-215	T	C
290	12-842-447	A	G
297	10-336-137	T	A
313	12-347-308	G	A
320	10-308-116	C	T
326	10-65-276	G	A
327	10-67-42	A	T
328	10-67-340	T	C
331	10-59-176	C	T
332	10-60-114	A	G
334	10-28-242	G	A
336	10-181-42	C	T
337	10-181-372	C	T
338	10-183-260	C	G
341	10-479-266	G	A
364	12-870-491	A	G
402	12-168-84	A	C
404	12-170-299	G	A
405	12-171-360	C	T
406	12-173-59	A	G
410	10-128-45	T	C
411	10-128-63	A	G
412	10-123-177	G	A
414	10-120-137	G	A
415	10-120-141	C	A
425	10-46-372	C	T
429	10-48-184	C	T
430	10-48-381	C	T
431	10-49-33	C	T
432	10-39-148	A	G
435	10-42-354	T	C
436	10-154-42	C	T
437	10-154-156	C	T
438	10-154-226	G	A
439	12-776-259	A	G
440	10-384-109	C	T
441	12-296-388	A	G
457	12-309-405	A	G
467	12-45-145	A	G
468	12-45-166	G	A
487	10-1-139	G	T

FIG.4C

488	10-1-212	G	T
489	10-1-241	C	A
491	10-9-185	T	C
492	10-9-264	C	G
493	10-11-22	T	C
495	10-13-256	C	T
496	10-13-282	T	C
497	10-15-281	T	G
498	10-17-142	C	T
499	10-18-302	C	T
500	10-23-331	G	A
501	10-25-152	T	C
502	10-25-258	C	T
503	10-3-103	C	T
504	10-3-144	T	C
505	10-3-275	G	T
506	10-5-227	A	C
507	10-7-155	T	C
508	10-7-383	C	T
509	10-7-98	G	C
533	10-20-274	A	G
534	10-24-90	A	C
536	10-24-221	G	T
546	10-8-92	T	C
547	10-8-94	C	T
548	12-61-472	C	T
560	12-76-238	G	T
566	10-176-85	C	T
567	10-176-51	C	T
568	10-176-207	G	T
569	10-176-397	C	A
570	10-177-219	A	C
575	10-234-179	AA	-
576	10-235-272	T	-
584	10-207-410	-	C
585	10-171-254	GG	-
586	12-94-110	-	AATT
588	10-55-115	TTATA	-
591	12-882-40	A	-
594	12-283-386	T	-
598	10-485-256	A	G
599	10-485-257	T	C
600	10-474-320	Insertion A	-
601	10-387-371	T	C
604	12-570-393	C	T
605	12-570-421	T	G
606	12-570-62	Insertion TG	-
607	10-4-144	C	A
608	10-4-161	A	C
609	10-4-270	G	C
610	10-4-340	A	G
611	10-4-369	C	T
612	10-4-420	G	T
613	10-13-396	Insertion AAT	-

FIG.4D

614	10-509-284	C	T
616	10-339-124	C	T
617	10-340-112	C	A
618	10-340-130	T	A
622	10-343-231	Insertion C	-
623	10-343-278	C	T
624	10-346-141	G	A
628	10-349-216	Insertion CTG	-
631	10-507-170	A	G
639	10-511-337	Deletion	-
640	10-512-36	G	C
641	10-511-62	C	T
642	10-512-318	G	A
643	10-513-250	G	A
644	10-513-262	T	C
645	10-513-352	G	A
646	10-513-365	G	A

FIG.5

SEQ ID NO.	BIALLELIC MARKER ID	1ST ALLELE	2ND ALLELE
22	10-498-192	A	G
31	10-347-203	A	G
32	10-347-220	A	G
37	10-349-97	A	G
39	10-349-224	G	T
42	10-341-116	A	G
81	10-249-185	A	G
123	10-397-201	G	T
189	10-371-257	A	C
230	10-93-133	C	T
244	10-106-288	C	T
245	10-106-378	C	T
296	10-336-58	C	T
323	10-306-265	A	G
324	10-52-386	C	T
325	10-62-240	G	C
329	10-55-265	C	T
330	10-57-278	C	T
333	10-27-176	A	G
339	10-475-163	A	G
344	10-482-145	A	G
413	10-123-402	A	G
416	10-179-39	C	T
417	10-180-65	G	C
418	10-179-257	G	T
426	10-46-36	T	A
521	10-295-201	G	T
528	10-10-328	G	A
529	10-12-52	C	T
530	10-14-46	C	T
532	10-20-111	A	C
535	10-24-204	A	G
537	10-24-234	A	G
538	10-24-288	A	G
539	10-24-311	G	C
541	10-8-39	A	C
542	10-8-120	A	G
543	10-8-154	G	C
544	10-8-101	A	T
545	10-8-86	C	T
596	10-370-132	C	T
597	10-370-254	C	T
615	10-509-295	Insertion and Deletion	

FIG.6A

SEQ ID NO.	POSITION RANGE OF NOVEL SEQUENCE
26	[569-588],[815-956]
29	[1-97],[203-450],[747-956]
30	[1-43],[149-396],[693-956]
31	[111-358],[655-956]
32	[94-341],[638-956]
33	[44-291],[588-956]
34	[1-214],[511-844]
36	[734-843]
37	[684-793]
38	[639-748]
39	[557-666]
41	[217-319],[721-781]
42	[1-96],[276-387],[881-956]
43	[72-184],[678-820]
46	[1-56],[193-400]
47	[855-874]
49	[1-101],[259-305]
50	[1-101],[259-305]
51	[1-101],[259-305]
79	[1-311],[512-1001]
80	[709-1001]
81	[1-231],[723-741]
83	[291-476]
95	[1-138],[532-662],[970-1001]
96	[59-254],[648-778],[918-1001]
97	[318-757]
106	[88-182],[309-461],[798-843]
107	[1-119],[246-398],[735-780]
123	[1-449],[568-1001]
162	[264-407],[801-833]
163	[249-392],[786-818]
175	[1-419]
185	[267-360],[549-599],[651-807],[851-1001]
186	[1-459],[691-1001]
187	[1-311],[557-1001]
188	[1-155],[662-1001]
189	[1-39],[554-1001]
251	[746-1001]
265	[315-445],[873-1001]
266	[174-304],[732-1001]
271	[1-173],[572-844],[884-917]
272	[1-75],[556-576]
289	[1-191]
290	[160-421]
296	[1-151],[681-1001]
297	[1-72],[602-1001]
313	[1-319]
320	[592-1001]
339	[1-24],[804-1001]
341	[1-364],[530-1001]
364	[1-270],[554-1001]
402	[1-319]
404	[1-319],[767-830]
405	[1-222],[639-1001]
439	[1-73],[608-900]

FIG.6B

440	[1-40],[732-1001]
441	[182-281]
457	[1-315],[838-1001]
467	[1-222],[647-1001]
468	[1-201],[626-954]
521	[1-138],[281-412],[529-880]
548	[60-80]
560	[539-810]
588	[406-418]
591	[150-320],[777-824],[864-1002]
594	[300-450]
596	[196-237],[920-1001]
597	[74-115],[798-1001]
598	[557-1001]
599	[556-1001]
600	[256-267],[669-670],[833-835]
604	[1-100],[540-650],[776-825],[878-881],[969-985]
605	[1-128],[568-678],[804-853],[906-909],[997-1001]
606	[210-320],[446-495],[548-551],[771-1001]
607	[1-54]
608	[1-54]
609	[1-54]
610	[1-54]
611	[1-54]
612	[1-54]
613	[1-30],[138-179]
614	[725-814]
615	[714-803]
616	[1-252],[634-713],[890-1001]
617	[106-155],[357-436],[613-727]
618	[88-137],[339-418],[595-709]
622	[138-178],[592-638],[863-1001]
623	[91-131],[545-591],[816-1001]
624	[1-29],[118-159],[190-232],[363-477],[581-830]
628	[587-698]
631	[199-459],[591-613],[703-803],[862-1000]
641	[1-159],[190-325]
646	[1-20]

FIG.7A

SEQ ID NO.	POSITION RANGE OF MICROSEQUENCING PRIMERS	COMPLEMENTARY POSITION RANGE OF MICROSEQUENCING PRIMERS
1	458-477	479-498
2	459-477*	479-498
3	458-477	479-498
4	458-477	479-498
5	458-477	479-498
6	458-477	479-498
7	458-477	479-498
8	458-477	479-498
9	458-477	479-498
10	458-477	479-498
11	458-477	479-497*
12	459-477*	479-498
13	459-477*	479-498
14	458-477	479-498
15	459-477*	479-498
16	459-477*	479-498
17	459-477*	479-498
18	458-477	479-498
19	459-477*	479-498
20	459-477*	479-498
21	458-477	479-497*
22	458-477	479-498
23	458-477	479-497*
24	458-477	479-498
25	459-477*	479-498
26	459-477*	479-498
27	458-477	479-498
28	458-477	479-498
29	458-477	479-498
30	458-477	479-498
31	459-477*	479-498
32	458-477	479-497*
33	458-477	479-498
34	458-477	479-498
35	458-477	479-498
36	458-477	479-498
37	458-477	479-497*
38	458-477	479-498
39	458-477	479-497*
40	458-477	479-498
41	458-477	479-498
42	458-477	479-497*
43	458-477	479-498
44	100-118*	120-139
45	224-242*	244-263
46	108-127	129-148
47	458-477	479-498
48	16-34*	36-55
49	110-128*	130-149
50	131-150	152-171
51	338-357	359-378
52	446-465	467-486
53	398-417	419-437*
54	209-228	230-248*
55	235-254	256-275

FIG.7B

56	458-477	479-498
57	282-301	303-322
58	159-178	180-199
59	188-206*	208-227
60	521-539*	541-560
61	147-165*	167-185*
62	435-454	456-475
63	312-331	333-352
64	331-350	352-371
65	344-363	365-384
66	79-98	100-119
67	127-146	148-167
68	245-264	266-285
69	480-499	501-519*
70	481-500	502-520*
71	482-500*	502-521
72	481-500	502-521
73	482-500*	502-521
74	481-500	502-521
75	481-500	502-521
76	481-500	502-521
77	481-500	502-521
78	481-500	502-521
79	481-500	502-520*
80	481-500	502-521
81	481-500	502-521
82	481-500	502-521
83	481-500	502-521
84	482-500*	502-521
85	480-499	501-520
86	481-500	502-521
87	481-500	502-521
88	481-500	502-521
89	481-500	502-521
90	481-500	502-521
91	482-500*	502-521
92	481-500	502-521
93	481-500	502-521
94	481-500	502-521
95	481-500	502-521
96	481-500	502-521
97	238-257	259-278
98	481-500	502-521
99	423-442	444-462*
100	293-312	314-333
101	481-500	502-520*
102	481-500	502-521
103	481-500	502-521
104	481-500	502-521
105	365-384	386-405
106	482-500*	502-521
107	481-500	502-521
108	360-379	381-400
109	303-322	324-343
110	297-316	318-337
111	481-500	502-521
112	481-500	502-521

FIG.7C

113	481-500	502-521
114	481-500	502-521
115	482-500*	502-521
116	209-228	230-249
117	466-485	487-506
118	481-500	502-521
119	481-500	502-521
120	481-500	502-521
121	481-500	502-521
122	481-500	502-521
123	481-500	502-521
124	481-500	502-521
125	481-500	502-521
126	477-496	498-517
127	481-500	502-521
128	481-500	502-521
129	481-500	502-521
130	481-500	502-521
131	481-500	502-521
132	481-500	502-521
133	481-500	502-521
134	481-500	502-521
135	481-500	502-521
136	481-500	502-521
137	481-500	502-521
138	481-500	502-521
139	481-500	502-521
140	481-500	502-521
141	481-500	502-521
142	481-500	502-520*
143	481-500	502-521
144	481-500	502-521
145	481-500	502-521
146	481-500	502-521
147	482-500*	502-521
148	481-500	502-521
149	481-500	502-520*
150	482-500*	502-521
151	481-500	502-521
152	481-500	502-521
153	481-500	502-521
154	481-500	502-521
155	481-500	502-521
156	481-500	502-521
157	481-500	502-521
158	481-500	502-521
159	481-500	502-521
160	481-500	502-521
161	481-500	502-521
162	481-500	502-521
163	481-500	502-521
164	488-507	509-528
165	481-500	502-521
166	481-500	502-521
167	481-500	502-521
168	481-500	502-521
169	481-500	502-521

FIG.7D

170	481-500	502-521
171	481-500	502-521
172	481-500	502-521
173	481-500	502-521
174	481-500	502-521
175	481-500	502-521
176	481-500	502-521
177	481-500	502-521
178	481-500	502-521
179	521-540	542-561
180	481-500	502-521
181	481-500	502-521
182	481-500	502-521
183	481-500	502-521
184	481-500	502-521
185	481-500	502-521
186	481-500	502-521
187	481-500	502-521
188	481-500	502-521
189	481-500	502-521
190	481-500	502-521
191	481-500	502-521
192	481-500	502-521
193	481-500	502-520*
194	481-500	502-521
195	481-500	502-520*
196	482-500*	502-521
197	481-500	502-521
198	481-500	502-520*
199	481-500	502-521
200	481-500	502-521
201	482-500*	502-521
202	481-500	502-521
203	481-500	502-520*
204	481-500	502-521
205	481-500	502-521
206	481-500	502-521
207	481-500	502-521
208	481-500	502-521
209	481-500	502-521
210	481-500	502-521
211	481-500	502-521
212	481-500	502-521
213	481-500	502-521
214	481-500	502-521
215	481-500	502-521
216	481-500	502-521
217	481-500	502-521
218	481-500	502-521
219	481-500	502-521
220	481-500	502-521
221	481-500	502-521
222	52-71	73-92
223	53-72	74-93
224	59-78	80-99
225	118-137	139-158
226	62-80*	82-101

FIG.7E

227	21-40	42-61
228	15-34	36-54*
229	254-273	275-294
230	113-132	134-153
231	178-196*	198-217
232	178-197	199-218
233	342-361	363-382
234	368-386*	388-407
235	390-408*	410-429
236	287-306	308-326*
237	227-246	248-267
238	274-293	295-313*
239	328-346*	348-367
240	84-103	105-124
241	304-322*	324-343
242	383-402	404-423
243	78-97	99-118
244	269-287*	289-308
245	361-379*	381-400
246	141-159*	161-180
247	187-205*	207-226
248	263-282	284-303
249	297-316	318-337
250	481-500	502-520*
251	481-500	502-521
252	482-500*	502-521
253	481-500	502-521
254	481-500	502-521
255	501-520	522-541
256	482-500*	502-521
257	481-500	502-521
258	482-500*	502-521
259	482-500*	502-521
260	481-500	502-521
261	481-500	502-521
262	481-500	502-521
263	480-499	501-520
264	481-500	502-521
265	481-500	502-521
266	481-500	502-521
267	481-500	502-521
268	481-500	502-521
269	481-500	502-521
270	481-500	502-521
271	481-500	502-521
272	481-500	502-521
273	481-500	502-521
274	481-500	502-521
275	482-501	503-522
276	481-500	502-521
277	463-482	484-503
278	481-500	502-521
279	481-500	502-521
280	480-499	501-520
281	456-475	477-496
282	478-497	499-518
283	481-500	502-521

FIG.7F

284	481-500	502-521
285	481-500	502-521
286	481-500	502-521
287	481-500	502-521
288	425-444	446-465
289	481-500	502-521
290	479-498	500-519
291	481-500	502-521
292	481-500	502-521
293	481-500	502-521
294	481-500	502-521
295	481-500	502-521
296	481-500	502-521
297	481-500	502-521
298	481-500	502-521
299	359-378	380-399
300	366-385	387-406
301	481-500	502-521
302	482-500*	502-521
303	481-500	502-521
304	481-500	502-520*
305	419-438	440-459
306	258-277	279-297*
307	306-325	327-346
308	318-337	339-358
309	393-412	414-433
310	482-500*	502-521
311	481-500	502-520*
312	482-500*	502-521
313	481-500	502-521
314	482-500*	502-521
315	481-500	502-520*
316	481-500	502-520*
317	482-500*	502-521
318	482-500*	502-521
319	482-500*	502-521
320	481-500	502-521
321	481-500	502-521
322	481-500	502-521
323	481-500	502-521
324	366-385	387-406
325	221-239*	241-260
326	256-275	277-295*
327	22-41	43-62
328	322-340*	342-361
329	245-263*	265-284
330	259-277*	279-298
331	157-175*	177-196
332	94-113	115-134
333	156-175	177-195*
334	222-241	243-261*
335	330-349	351-369*
336	23-41*	43-62
337	355-373*	375-394
338	239-258	260-279
339	481-500	502-521
340	482-500*	502-521

FIG.7G

341	481-500	502-521
342	481-500	502-521
343	481-500	502-521
344	481-500	502-521
345	481-500	502-521
346	481-500	502-521
347	481-500	502-521
348	481-500	502-521
349	481-500	502-521
350	456-475	477-496
351	482-500*	502-521
352	481-500	502-521
353	481-500	502-521
354	481-500	502-521
355	481-500	502-521
356	481-500	502-521
357	481-500	502-521
358	482-500*	502-521
359	481-500	502-521
360	481-500	502-521
361	481-500	502-521
362	481-500	502-521
363	481-500	502-521
364	481-500	502-521
365	481-500	502-521
366	481-500	502-521
367	481-500	502-521
368	481-500	502-521
369	482-500*	502-521
370	481-500	502-521
371	481-500	502-521
372	481-500	502-521
373	481-500	502-521
374	481-500	502-521
375	481-500	502-521
376	481-500	502-521
377	481-500	502-521
378	481-500	502-521
379	481-500	502-521
380	481-500	502-521
381	481-500	502-521
382	481-500	502-521
383	481-500	502-521
384	481-500	502-521
385	481-500	502-521
386	481-500	502-521
387	482-500*	502-521
388	481-500	502-521
389	481-500	502-521
390	481-500	502-521
391	481-500	502-521
392	482-500*	502-521
393	481-500	502-521
394	459-478	480-499
395	481-500	502-521
396	481-500	502-520*
397	481-500	502-521

FIG.7H

398	481-500	502-521
399	481-500	502-521
400	508-527	529-548
401	481-500	502-520*
402	482-500*	502-521
403	481-500	502-521
404	481-500	502-520*
405	482-500*	502-521
406	481-500	502-520*
407	481-500	502-520*
408	481-500	502-520*
409	481-500	502-521
410	26-44*	46-65
411	43-62	64-83
412	157-176	178-197
413	382-401	403-422
414	117-135*	137-156
415	121-139*	141-160
416	19-38	40-59
417	46-64*	66-85
418	237-256	258-276*
425	350-368*	370-389
426	15-34	36-55
427	82-101	103-122
428	105-123*	125-144
429	164-182*	184-203
430	362-381	383-402
431	14-32*	34-53
432	130-149	151-169*
433	202-221	223-241*
434	230-249	251-270
435	334-353	355-373*
436	22-41	43-62
437	137-155*	157-176
438	206-225	227-246
439	481-500	502-521
440	481-500	502-521
441	482-500*	502-521
442	481-500	502-521
443	481-500	502-521
444	481-500	502-521
445	481-500	502-521
446	482-500*	502-521
447	482-500*	502-521
448	481-500	502-521
449	481-500	502-521
450	481-500	502-521
451	481-500	502-521
452	481-500	502-521
453	481-500	502-521
454	481-500	502-521
455	481-500	502-521
456	481-500	502-521
457	481-500	502-521
458	481-500	502-521
459	482-500*	502-521
460	481-500	502-520*

FIG.7I

461	481-500	502-521
462	481-500	502-521
463	482-500*	502-521
464	481-500	502-521
465	482-500*	502-521
466	481-500	502-521
467	481-500	502-521
468	481-500	502-521
469	481-500	502-521
470	481-500	502-520*
471	482-500*	502-521
472	481-500	502-521
473	481-500	502-521
474	481-500	502-521
475	481-500	502-521
476	481-500	502-521
477	482-500*	502-521
478	481-500	502-521
479	481-500	502-521
480	483-500*	502-521
481	481-500	502-521
482	481-500	502-521
483	481-500	502-521
484	481-500	502-520*
485	482-500*	502-520*
486	481-500	502-521
487	119-138	140-158*
488	192-211	213-231*
489	222-240*	242-261
491	166-184*	186-205
492	245-263*	265-284
493	2-21	23-42
494	133-151*	153-172
495	237-255*	257-276
496	263-281*	283-302
497	261-280	282-301
498	123-141*	143-162
499	282-301	303-322
500	311-330	332-351
501	133-151*	153-172
502	238-257	259-278
503	84-102*	104-123
504	125-143*	145-164
505	255-274	276-294*
506	208-226*	228-247
507	136-154*	156-175
508	361-380	382-401
509	78-97	99-118
510	481-500	502-520*
511	482-500*	502-521
512	481-500	502-520*
513	481-500	502-521
514	481-500	502-521
515	481-500	502-521
516	481-500	502-521
517	482-500*	502-521
518	482-500*	502-521

FIG.7J

519	481-500	502-521
520	481-500	502-521
521	481-500	502-521
522	481-500	502-521
523	481-500	502-521
524	481-500	502-521
525	481-500	502-521
526	481-500	502-521
527	481-500	502-521
528	307-326	328-346*
529	32-51	53-72
530	27-45*	47-66
532	91-109*	111-130
533	253-272	274-292*
534	71-89*	91-110
535	184-203	205-223*
536	201-220	222-241
537	214-233	235-253*
538	268-287	289-308
539	291-310	312-331
541	20-38*	40-59
542	100-119	121-140
543	134-153	155-174
544	81-100	102-119*
545	67-85*	87-106
546	72-91	93-112
547	74-93	95-114
548	481-500	502-521
549	396-415	417-436
550	68-87	89-108
551	268-286*	288-307
552	419-438	440-459
553	482-500*	502-521
554	482-500*	502-521
555	481-500	502-520*
556	481-500	502-521
557	482-500*	502-521
558	481-500	502-521
559	481-500	502-520*
560	481-500	502-520*
561	481-500	502-521
562	481-500	502-520*
563	481-500	502-521
564	481-500	502-520*
565	230-249	251-270
566	66-84*	86-105
567	31-50	52-71
568	187-206	208-227
569	378-396*	398-416*
570	200-218*	220-239
571	65-84	-
572	251-270	-
573	143-162	-
574	62-81	-
575	195-213*	-
576	471-490	-
577	478-497	-

FIG.7K

578	477-496	-
579	478-497	-
580	-	459-477*
581	-	498-516*
582	478-497	-
583	477-496	-
584	389-408	-
585	235-254	-
586	479-497 *	-
587	478-497	-
588	95-113 *	-
589	478-497	-
590	478-497	-
591	478-497	-
592	478-497	-
593	479-498	-
594	478-497	-
595	438-457	-
596	481-500	502-521
597	481-500	502-521
598	481-500	502-521
599	481-500	502-521
600	481-500	-
601	481-500	502-521
602	481-500	502-521
603	481-500	502-521
604	481-500	502-521
605	481-500	502-521
606	-	502-521
607	121-140	142-161
608	138-157	159-178
609	247-266	268-287
610	317-336	338-357
611	346-365	367-386
612	397-416	418-437
613	374-395	-
614	481-500	502-521
615	481-500	-
616	481-500	502-521
617	481-500	502-521
618	481-500	502-521
619	481-500	502-521
620	481-500	-
621	481-500	502-521
622	481-500	-
623	481-500	502-521
624	481-500	502-520*
625	480-499	501-520
626	481-500	502-521
627	481-500	502-521
628	481-500	-

FIG.7L

629	481-500	502-521
630	481-500	502-521
631	481-500	502-521
632	481-500	502-521
633	481-500	502-521
634	481-500	502-521
635	481-500	502-521
636	481-500	502-521
637	481-500	502-521
638	481-500	-
639	481-500	-
640	481-500	502-521
641	481-500	502-521
642	481-500	502-521
643	481-500	502-521
644	481-500	502-521
645	481-500	502-521
646	481-500	502-521
647	481-500	502-521
648	481-500	502-521
649	481-500	502-521
650	481-500	502-521

FIG.8A

SEQ ID NO.	POSITION RANGE OF AMPLIFICATION PRIMERS	COMPLEMENTARY POSITION RANGE OF AMPLIFICATION PRIMERS
1	361-379	761-780
2	181-199	581-600
3	164-182	564-583
4	324-343	536-553
5	294-310	743-760
6	221-237	670-687
7	70-86	519-536
8	326-343	760-780
9	307-324	557-575
10	237-254	487-505
11	153-170	590-607
12	121-139	522-541
13	304-322	705-723
14	268-286	669-687
15	245-263	646-664
16	209-227	610-628
17	152-170	553-571
18	189-206	525-542
19	120-137	526-543
20	88-105	494-511
21	315-334	741-760
22	287-306	621-638
23	266-286	764-782
24	271-291	769-787
25	238-257	617-637
26	222-239	635-654
27	140-157	553-572
28	405-422	826-845
29	368-385	789-808
30	314-331	735-754
31	276-293	697-716
32	259-276	680-699
33	209-226	630-649
34	132-149	553-572
35	90-109	488-507
36	432-451	829-848
37	382-401	779-798
38	337-356	734-753
39	255-274	652-671
40	114-133	511-530
41	447-464	845-864
42	363-380	771-789
43	160-177	568-586
44	1-20	450-469
45	1-19	380-399
46	1-20	380-400
47	398-415	835-854
48	1-21	487-507
49	1-20	429-448
50	1-20	429-448
51	1-20	429-448
52	1-20	479-499
53	1-20	467-486

FIG.8B

54	1-20	465-485
55	1-20	465-485
56	76-96	505-525
57	1-21	387-407
58	1-20	449-468
59	1-20	449-468
60	1-19	581-598
61	1-19	490-508
62	1-19	490-508
63	1-21	437-456
64	1-21	437-456
65	1-20	401-420
66	1-19	470-490
67	1-19	470-490
68	1-19	470-490
69	478-495	879-898
70	119-137	540-557
71	183-200	603-620
72	138-157	538-556
73	472-490	900-917
74	431-449	859-876
75	374-392	802-819
76	243-261	671-688
77	113-131	541-558
78	161-179	561-580
79	220-238	620-638
80	261-279	595-614
81	317-335	720-738
82	374-393	732-751
83	293-312	701-720
84	470-488	901-921
85	185-203	590-609
86	90-107	509-528
87	83-103	652-671
88	140-160	709-728
89	361-381	930-949
90	437-456	885-905
91	298-318	727-747
92	385-404	832-852
93	332-351	779-799
94	234-253	681-701
95	347-367	816-836
96	463-483	932-952
97	136-155	591-611
98	75-94	530-550
99	290-310	669-688
100	290-310	669-688
101	325-343	756-776
102	71-90	503-519
103	136-154	571-590
104	417-435	930-949
105	5-23	530-549
106	285-304	749-769
107	222-241	686-706

FIG.8C

108	4-22	446-466
109	198-218	634-654
110	53-70	504-523
111	450-468	881-899
112	93-111	524-542
113	285-303	746-766
114	103-121	564-584
115	147-165	587-607
116	9-29	435-455
117	187-206	636-656
118	406-425	797-816
119	384-403	775-794
120	354-373	745-764
121	216-235	607-626
122	402-420	735-752
123	301-319	707-724
124	324-341	618-635
125	134-153	533-552
126	478-495	906-924
127	399-416	827-845
128	178-195	606-624
129	475-493	879-896
130	179-197	583-600
131	265-284	694-714
132	277-295	720-740
133	360-378	867-887
134	309-327	816-836
135	442-462	898-918
136	383-403	839-859
137	365-385	821-841
138	356-376	812-832
139	104-124	560-580
140	389-408	879-898
141	338-357	828-847
142	337-356	827-846
143	57-76	547-566
144	301-320	682-701
145	10-30	524-542
146	386-406	844-862
147	47-67	547-566
148	268-288	768-787
149	478-498	903-922
150	391-410	828-846
151	385-404	822-840
152	316-336	745-765
153	378-398	807-827
154	427-447	856-876
155	67-86	512-530
156	112-131	557-575
157	380-399	825-843
158	387-406	832-850
159	183-201	672-692
160	321-340	767-787
161	200-219	646-666

FIG.8D

162	262-281	761-780
163	247-266	746-765
164	1-21	527-544
165	42-62	568-585
166	153-173	679-696
167	441-461	967-984
168	364-382	900-920
169	127-145	663-683
170	330-349	769-788
171	284-303	723-742
172	176-195	615-634
173	110-129	549-568
174	222-242	688-708
175	118-135	628-647
176	245-264	749-768
177	128-147	632-651
178	79-98	583-602
179	1-21	545-563
180	116-136	660-678
181	355-373	811-828
182	403-421	859-876
183	74-94	649-668
184	408-428	983-1002
185	224-242	627-646
186	366-385	765-784
187	221-239	646-665
188	111-129	512-531
189	246-263	661-679
190	114-133	610-627
191	9-28	505-522
192	107-126	537-557
193	405-424	891-911
194	216-234	653-673
195	180-198	617-637
196	129-149	661-678
197	220-240	752-769
198	352-372	849-867
199	349-369	856-875
200	461-481	968-987
201	369-389	798-817
202	110-130	539-558
203	329-346	723-741
204	62-79	567-587
205	130-150	558-578
206	131-151	559-579
207	267-284	795-815
208	304-324	770-787
209	216-234	712-731
210	188-207	659-674
211	172-192	645-663
212	194-214	674-693
213	63-83	543-562
214	236-256	687-705
215	268-288	719-737

FIG.8E

216	335-355	786-804
217	463-482	899-918
218	72-91	508-527
219	115-135	555-573
220	236-256	681-700
221	290-310	735-754
222	1-18	345-362
223	1-18	345-362
224	1-18	345-362
225	1-18	345-362
226	1-18	349-368
227	1-19	401-420
228	1-18	409-427
229	1-18	402-420
230	1-19	407-426
231	1-19	403-420
232	1-19	403-420
233	1-18	363-380
234	1-12	444-464
235	1-12	444-464
236	1-18	343-361
237	1-19	418-435
238	1-19	418-435
239	1-19	418-435
240	1-19	420-439
241	1-19	420-439
242	1-19	420-439
243	1-19	406-424
244	1-19	406-424
245	1-19	406-424
246	1-18	354-371
247	1-18	354-371
248	1-18	354-371
249	1-20	408-425
250	130-149	562-579
251	444-464	875-894
252	133-151	564-584
253	411-429	842-862
254	292-312	819-839
255	6-26	533-553
256	101-121	547-564
257	419-437	866-886
258	347-366	776-796
259	253-273	749-768
260	127-146	656-675
261	408-428	849-859
262	119-139	560-570
263	118-138	559-569
264	111-131	552-562
265	288-307	791-810
266	147-166	650-669
267	65-84	527-546
268	17-37	542-562
269	438-458	963-983

FIG.8F

270	329-347	774-794
271	337-357	792-811
272	92-112	540-559
273	315-335	746-764
274	330-350	761-779
275	331-351	762-780
276	424-444	855-873
277	306-326	737-757
278	449-468	879-898
279	368-386	909-929
280	266-284	807-827
281	242-260	783-803
282	242-260	783-803
283	227-245	768-788
284	323-342	895-915
285	448-468	877-897
286	98-118	527-547
287	128-148	557-577
288	1-21	483-502
289	248-268	697-715
290	478-498	927-945
291	245-265	649-667
292	415-435	843-863
293	294-312	786-796
294	379-397	835-855
295	293-312	723-742
296	444-462	845-863
297	365-383	766-784
298	270-288	671-689
299	276-296	733-753
300	276-296	733-753
301	227-247	684-704
302	301-318	733-751
303	289-306	721-739
304	373-393	831-851
305	5-23	442-462
306	130-148	577-597
307	130-148	577-597
308	130-148	577-597
309	130-148	577-597
310	288-307	756-775
311	445-463	901-921
312	41-61	509-526
313	195-213	636-656
314	307-327	816-834
315	362-381	794-814
316	182-200	677-696
317	114-132	609-628
318	252-270	697-717
319	262-280	741-761
320	386-403	754-771
321	228-245	632-651
322	426-443	827-846
323	238-255	660-678

FIG.8G

324	1-18	412-429
325	1-20	419-438
326	1-19	408-425
327	2-20	403-422
328	2-20	403-422
329	1-17	405-418
330	1-19	410-429
331	1-19	401-420
332	1-20	409-428
333	1-18	290-307
334	3-20	410-429
335	1-18	368-385
336	1-18	424-443
337	1-18	424-443
338	1-18	430-447
339	340-358	739-758
340	299-317	817-837
341	236-253	638-657
342	152-169	554-573
343	108-125	510-529
344	357-374	779-798
345	438-457	942-962
346	30-49	534-554
347	234-254	674-694
348	328-348	768-788
349	463-483	903-923
350	452-471	965-985
351	371-391	827-846
352	468-488	924-943
353	367-387	866-885
354	232-251	668-688
355	153-171	607-626
356	137-155	591-610
357	50-68	504-523
358	79-98	589-609
359	455-474	898-918
360	322-340	789-809
361	305-323	772-792
362	221-239	688-708
363	374-394	813-833
364	11-31	528-548
365	450-470	893-911
366	209-229	652-670
367	151-169	667-685
368	285-303	801-819
369	251-271	727-747
370	242-261	745-765
371	257-276	760-780
372	431-450	934-954
373	443-462	875-895
374	433-452	865-885
375	423-442	855-875
376	349-369	839-859
377	83-103	573-593

FIG.8H

378	435-453	951-971
379	64-82	580-600
380	292-311	723-743
381	113-132	544-564
382	229-248	674-694
383	235-255	677-696
384	372-392	814-833
385	446-466	888-907
386	449-469	891-910
387	246-266	674-694
388	399-419	827-847
389	184-204	683-701
390	450-470	949-967
391	148-168	578-598
392	253-273	683-703
393	365-385	795-815
394	426-445	976-995
395	309-329	745-765
396	348-368	870-889
397	138-158	622-640
398	106-126	595-615
399	216-236	705-725
400	6-26	558-576
401	116-134	600-619
402	105-125	566-584
403	386-406	847-865
404	296-315	781-799
405	142-160	521-541
406	443-460	901-921
407	287-306	765-785
408	217-236	664-683
409	399-418	846-865
410	1-20	349-366
411	1-20	349-366
412	1-18	412-429
413	1-18	412-429
414	1-16	333-349
415	1-16	333-349
416	1-18	400-420
417	1-18	407-424
418	1-18	400-420
425	1-18	398-417
426	1-18	398-417
427	1-20	404-423
428	1-20	404-423
429	1-18	403-421
430	1-18	403-421
431	1-19	402-419
432	1-20	405-422
433	1-18	335-352
434	1-18	335-352
435	1-18	413-432
436	1-19	283-300
437	1-19	283-300

FIG.8I

438	1-19	283-300
439	243-263	674-692
440	393-412	802-819
441	114-134	543-563
442	123-141	542-561
443	386-403	792-809
444	153-170	559-576
445	409-426	821-838
446	355-374	783-803
447	90-109	600-620
448	199-219	670-690
449	467-486	915-935
450	383-403	812-832
451	211-229	688-707
452	33-53	586-605
453	376-395	840-860
454	92-111	556-576
455	124-142	553-573
456	286-306	745-764
457	456-475	884-904
458	397-415	826-845
459	49-67	536-556
460	210-229	679-698
461	188-207	600-617
462	121-138	521-540
463	210-228	647-667
464	95-113	511-530
465	147-166	530-550
466	164-183	547-567
467	357-375	804-824
468	336-354	783-803
469	198-216	645-665
470	410-429	792-811
471	370-388	756-776
472	441-459	827-847
473	403-421	904-921
474	180-198	681-698
475	134-152	635-652
476	466-484	967-984
477	180-200	613-630
478	102-122	535-552
479	113-133	546-563
480	329-347	717-734
481	348-366	736-753
482	464-482	852-869
483	451-471	881-900
484	151-168	672-692
485	180-197	701-721
486	469-486	990-1010
487	1-20	429-448
488	1-20	429-448
489	1-20	429-448
491	1-18	332-351
492	1-18	332-351

FIG.8J

493	1-18	402-421
494	1-19	402-420
495	1-19	402-420
496	1-19	402-420
497	1-20	409-428
498	1-20	413-431
499	1-19	403-422
500	1-19	427-446
501	1-18	408-427
502	1-18	408-427
503	1-19	411-430
504	1-19	411-430
505	1-19	411-430
506	1-20	404-421
507	1-20	399-418
508	1-20	399-418
509	1-20	399-418
510	232-252	703-723
511	415-433	862-882
512	438-456	920-940
513	289-307	771-791
514	295-313	812-829
515	137-155	654-671
516	135-153	652-669
517	165-183	647-667
518	124-144	590-610
519	246-264	650-669
520	198-216	602-621
521	301-320	701-720
522	423-440	833-850
523	131-148	541-558
524	384-401	806-825
525	348-365	770-789
526	453-470	802-820
527	317-334	666-684
528	1-20	414-433
529	1-18	330-349
530	1-18	414-431
532	1-19	400-419
533	1-19	400-419
534	1-20	416-435
535	1-20	416-435
536	1-20	416-435
537	1-20	416-435
538	1-20	416-435
539	1-20	416-435
541	1-20	427-446
542	1-20	427-446
543	1-20	427-446
544	1-20	427-446
545	1-20	427-446
546	1-20	427-446
547	1-20	427-446
548	39-57	548-568

FIG.8K

549	15-35	445-464
550	15-35	445-464
551	19-38	407-425
552	342-360	777-797
553	355-374	784-804
554	105-124	534-554
555	187-206	703-721
556	353-370	799-818
557	453-470	899-918
558	446-463	892-911
559	88-106	519-538
560	264-283	671-689
561	288-306	772-792
562	28-46	512-532
563	101-121	595-614
564	220-240	714-733
565	1-21	436-454
566	1-18	404-423
567	1-18	404-423
568	1-18	404-423
569	1-18	404-423
570	1-18	400-419
571	1-20	429-448
572	1-20	479-499
573	1-21	387-407
574	1-19	581-598
575	36-56	346-366
576	220-237	553-571
577	157-176	515-534
578	140-159	531-550
579	441-459	948-968
580	133-153	647-665
581	476-494	907-927
582	425-445	939-957
583	292-310	788-807
584	1-18	443-463
585	1-19	333-351
586	389-409	916-936
587	207-227	638-658
588	1-19	401-418
589	376-395	889-909
590	324-344	767-785
591	459-478	904-924
592	281-301	711-731
593	147-166	657-677
594	114-132	547-567
595	235-254	618-638
596	371-389	772-791
597	249-267	650-669
598	246-264	652-669
599	245-263	651-668
600	182-199	590-609
601	131-149	535-552

This Page Blank (uspto)

FIG.8L

602	255-274	719-738
603	360-379	824-843
604	409-428	873-892
605	437-456	901-920
606	79-98	543-562
607	1-18	427-444
608	1-18	427-444
609	1-18	427-444
610	1-18	427-444
611	1-18	427-444
613	1-19	402-420
614	218-235	618-637
615	207-224	607-626
616	379-396	776-795
617	390-408	791-809
618	372-390	773-791
619	264-282	665-683
620	201-220	606-623
621	129-148	534-551
622	271-288	684-703
623	224-241	637-656
624	363-381	763-780
625	478-496	878-895
626	241-259	641-658
627	199-217	599-616
628	286-305	683-702
629	172-189	513-532
630	430-447	771-790
631	332-350	739-758
632	181-199	588-607
633	149-167	556-575
634	138-156	545-564
635	97-115	504-523
636	311-328	727-746
637	257-274	673-692
638	329-347	729-748
639	165-182	569-587
640	147-164	551-569
641	440-457	844-862
642	184-203	607-626
643	252-269	660-677
644	240-257	648-665
645	150-167	558-575
646	137-154	545-562
647	402-420	722-740
648	378-396	630-648
649	309-327	561-579
650	431-449	885-903

FIG.9A

SEQ ID NO.	POSITION RANGE OF PROBES
1	466-490
2	466-490
3	466-490
4	466-490
5	466-490
6	466-490
7	466-490
8	466-490
9	466-490
10	466-490
11	466-490
12	466-490
13	466-490
14	466-490
15	466-490
16	466-490
17	466-490
18	466-490
19	466-490
20	466-490
21	466-490
22	466-490
23	466-490
24	466-490
25	466-490
26	466-490
27	466-490
28	466-490
29	466-490
30	466-490
31	466-490
32	466-490
33	466-490
34	466-490
35	466-490
36	466-490
37	466-490
38	466-490
39	466-490
40	466-490
41	466-490
42	466-490
43	466-490
44	107-131
45	231-255
46	116-140
47	466-490
48	23-47
49	117-141
50	139-163
51	346-370
52	454-478
53	406-430
54	217-241
55	243-267
56	466-490
57	290-314
58	167-191
59	195-219
60	528-552
61	154-178
62	443-467
63	320-344
64	339-363
65	352-376
66	87-111
67	135-159
68	253-277
69	488-512
70	489-513
71	489-513
72	489-513
73	489-513
74	489-513
75	489-513
76	489-513
77	489-513
78	489-513
79	489-513
80	489-513
81	489-513
82	489-513
83	489-513
84	489-513
85	488-512
86	489-513
87	489-513
88	489-513
89	489-513
90	489-513
91	489-513
92	489-513
93	489-513
94	489-513
95	489-513
96	489-513
97	246-270
98	489-513
99	431-455
100	301-325
101	489-513
102	489-513
103	489-513
104	489-513
105	373-397
106	489-513
107	489-513
108	368-392
109	311-335
110	305-329
111	489-513
112	489-513

FIG.9B

113	489-513
114	489-513
115	489-513
116	217-241
117	474-498
118	489-513
119	489-513
120	489-513
121	489-513
122	489-513
123	489-513
124	489-513
125	489-513
126	485-509
127	489-513
128	489-513
129	489-513
130	489-513
131	489-513
132	489-513
133	489-513
134	489-513
135	489-513
136	489-513
137	489-513
138	489-513
139	489-513
140	489-513
141	489-513
142	489-513
143	489-513
144	489-513
145	489-513
146	489-513
147	489-513
148	489-513
149	489-513
150	489-513
151	489-513
152	489-513
153	489-513
154	489-513
155	489-513
156	489-513
157	489-513
158	489-513
159	489-513
160	489-513
161	489-513
162	489-513
163	489-513
164	496-520
165	489-513
166	489-513
167	489-513
168	489-513
169	489-513
170	489-513
171	489-513
172	489-513
173	489-513
174	489-513
175	489-513
176	489-513
177	489-513
178	489-513
179	529-553
180	489-513
181	489-513
182	489-513
183	489-513
184	489-513
185	489-513
186	489-513
187	489-513
188	489-513
189	489-513
190	489-513
191	489-513
192	489-513
193	489-513
194	489-513
195	489-513
196	489-513
197	489-513
198	489-513
199	489-513
200	489-513
201	489-513
202	489-513
203	489-513
204	489-513
205	489-513
206	489-513
207	489-513
208	489-513
209	489-513
210	489-513
211	489-513
212	489-513
213	489-513
214	489-513
215	489-513
216	489-513
217	489-513
218	489-513
219	489-513
220	489-513
221	489-513
222	60-84
223	61-85
224	67-91
225	126-150
226	69-93

FIG.9C

227	29-53
228	23-47
229	262-286
230	121-145
231	185-209
232	186-210
233	350-374
234	375-399
235	397-421
236	295-319
237	235-259
238	282-306
239	335-359
240	92-116
241	311-335
242	391-415
243	86-110
244	276-300
245	368-392
246	148-172
247	194-218
248	271-295
249	305-329
250	489-513
251	489-513
252	489-513
253	489-513
254	489-513
255	509-533
256	489-513
257	489-513
258	489-513
259	489-513
260	489-513
261	489-513
262	489-513
263	488-512
264	489-513
265	489-513
266	489-513
267	489-513
268	489-513
269	489-513
270	489-513
271	489-513
272	489-513
273	489-513
274	489-513
275	490-514
276	489-513
277	471-495
278	489-513
279	489-513
280	488-512
281	464-488
282	486-510
283	489-513

284	489-513
285	489-513
286	489-513
287	489-513
288	433-457
289	489-513
290	487-511
291	489-513
292	489-513
293	489-513
294	489-513
295	489-513
296	489-513
297	489-513
298	489-513
299	367-391
300	374-398
301	489-513
302	489-513
303	489-513
304	489-513
305	427-451
306	266-290
307	314-338
308	326-350
309	401-425
310	489-513
311	489-513
312	489-513
313	489-513
314	489-513
315	489-513
316	489-513
317	489-513
318	489-513
319	489-513
320	489-513
321	489-513
322	489-513
323	489-513
324	374-398
325	228-252
326	264-288
327	30-54
328	329-353
329	252-276
330	266-290
331	164-188
332	102-126
333	164-188
334	230-254
335	338-362
336	30-54
337	362-386
338	247-271
339	489-513
340	489-513

FIG.9D

341	489-513
342	489-513
343	489-513
344	489-513
345	489-513
346	489-513
347	489-513
348	489-513
349	489-513
350	464-488
351	489-513
352	489-513
353	489-513
354	489-513
355	489-513
356	489-513
357	489-513
358	489-513
359	489-513
360	489-513
361	489-513
362	489-513
363	489-513
364	489-513
365	489-513
366	489-513
367	489-513
368	489-513
369	489-513
370	489-513
371	489-513
372	489-513
373	489-513
374	489-513
375	489-513
376	489-513
377	489-513
378	489-513
379	489-513
380	489-513
381	489-513
382	489-513
383	489-513
384	489-513
385	489-513
386	489-513
387	489-513
388	489-513
389	489-513
390	489-513
391	489-513
392	489-513
393	489-513
394	467-491
395	489-513
396	489-513
397	489-513

398	489-513
399	489-513
400	516-540
401	489-513
402	489-513
403	489-513
404	489-513
405	489-513
406	489-513
407	489-513
408	489-513
409	489-513
410	33-57
411	51-75
412	165-189
413	390-414
414	124-148
415	128-152
416	27-51
417	53-77
418	245-269
425	357-381
426	23-47
427	90-114
428	112-136
429	171-195
430	370-394
431	21-45
432	138-162
433	210-234
434	238-262
435	342-366
436	30-54
437	144-168
438	214-238
439	489-513
440	489-513
441	489-513
442	489-513
443	489-513
444	489-513
445	489-513
446	489-513
447	489-513
448	489-513
449	489-513
450	489-513
451	489-513
452	489-513
453	489-513
454	489-513
455	489-513
456	489-513
457	489-513
458	489-513
459	489-513
460	489-513

FIG.9E

461	489-513
462	489-513
463	489-513
464	489-513
465	489-513
466	489-513
467	489-513
468	489-513
469	489-513
470	489-513
471	489-513
472	489-513
473	489-513
474	489-513
475	489-513
476	489-513
477	489-513
478	489-513
479	489-513
480	489-513
481	489-513
482	489-513
483	489-513
484	489-513
485	489-513
486	489-513
487	127-151
488	200-224
489	229-253
491	173-197
492	252-276
493	10-34
494	140-164
495	244-268
496	270-294
497	269-293
498	130-154
499	290-314
500	319-343
501	140-164
502	246-270
503	91-115
504	132-156
505	263-287
506	215-239
507	143-167
508	369-393
509	86-110
510	489-513
511	489-513
512	489-513
513	489-513
514	489-513
515	489-513
516	489-513
517	489-513
518	489-513

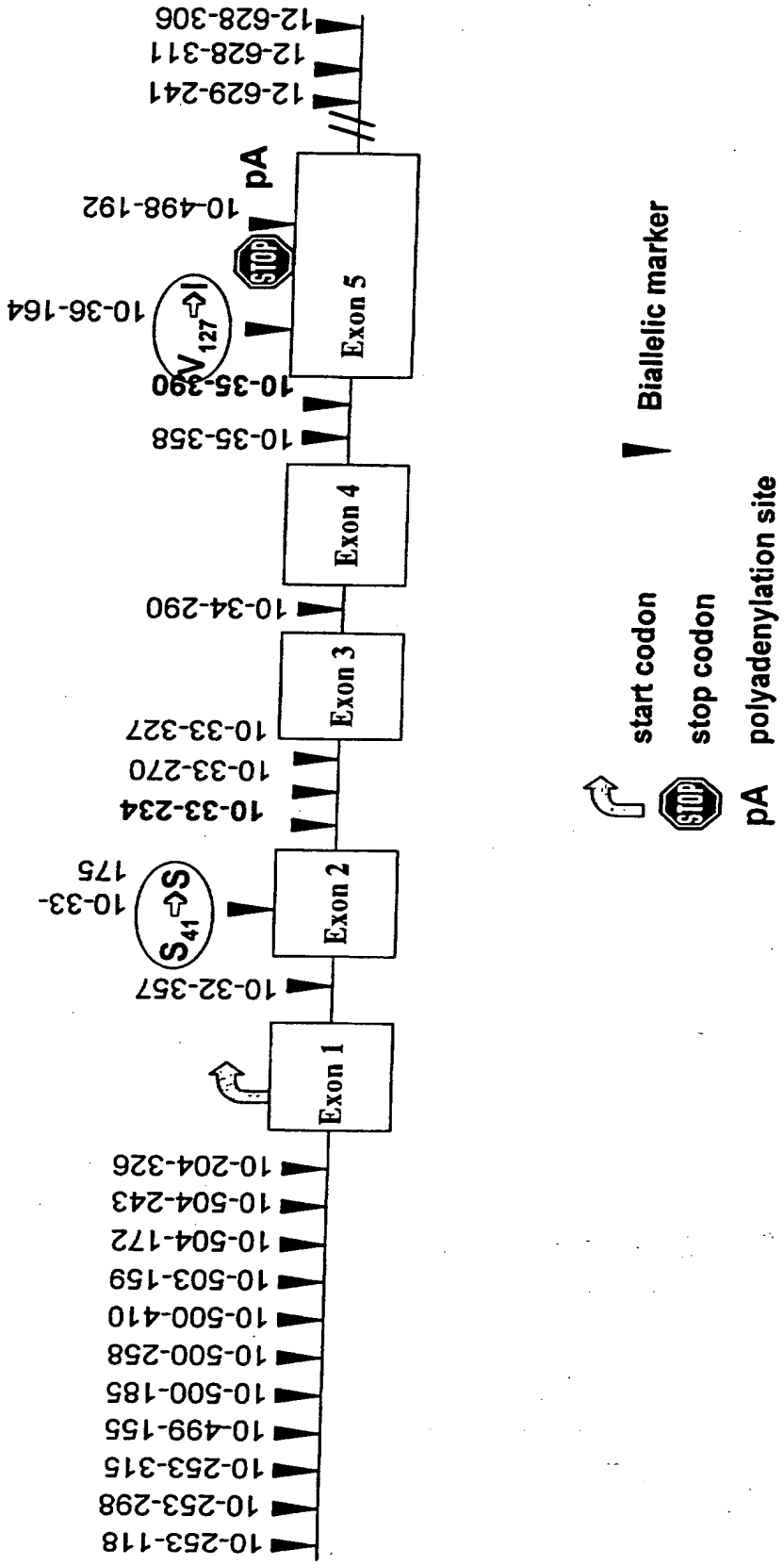
519	489-513
520	489-513
521	489-513
522	489-513
523	489-513
524	489-513
525	489-513
526	489-513
527	489-513
528	315-339
529	40-64
530	34-58
532	98-122
533	261-285
534	78-102
535	192-216
536	209-233
537	222-246
538	276-300
539	299-323
541	27-51
542	108-132
543	142-166
544	89-113
545	74-98
546	80-104
547	82-106
548	489-513
549	404-428
550	76-100
551	275-299
552	427-451
553	489-513
554	489-513
555	489-513
556	489-513
557	489-513
558	489-513
559	489-513
560	489-513
561	489-513
562	489-513
563	489-513
564	489-513
565	238-262
566	73-97
567	39-63
568	195-219
569	385-409
570	207-231
596	489-513
597	489-513
598	489-513
599	489-513
601	489-513
602	489-513

FIG.9F

603	489-513
604	489-513
605	489-513
607	129-153
608	146-170
609	255-279
610	325-349
611	354-378
612	405-429
614	489-513
616	489-513
617	489-513
618	489-513
619	489-513
621	489-513
623	489-513
624	489-513
625	488-512
626	489-513
627	489-513
629	489-513
630	489-513
631	489-513
632	489-513
633	489-513
634	489-513
635	489-513
636	489-513
637	489-513
640	489-513
641	489-513
642	489-513
643	489-513
644	489-513
645	489-513
646	489-513
647	489-513
648	489-513
649	489-513
650	489-513

FIG.10

FLAP gene



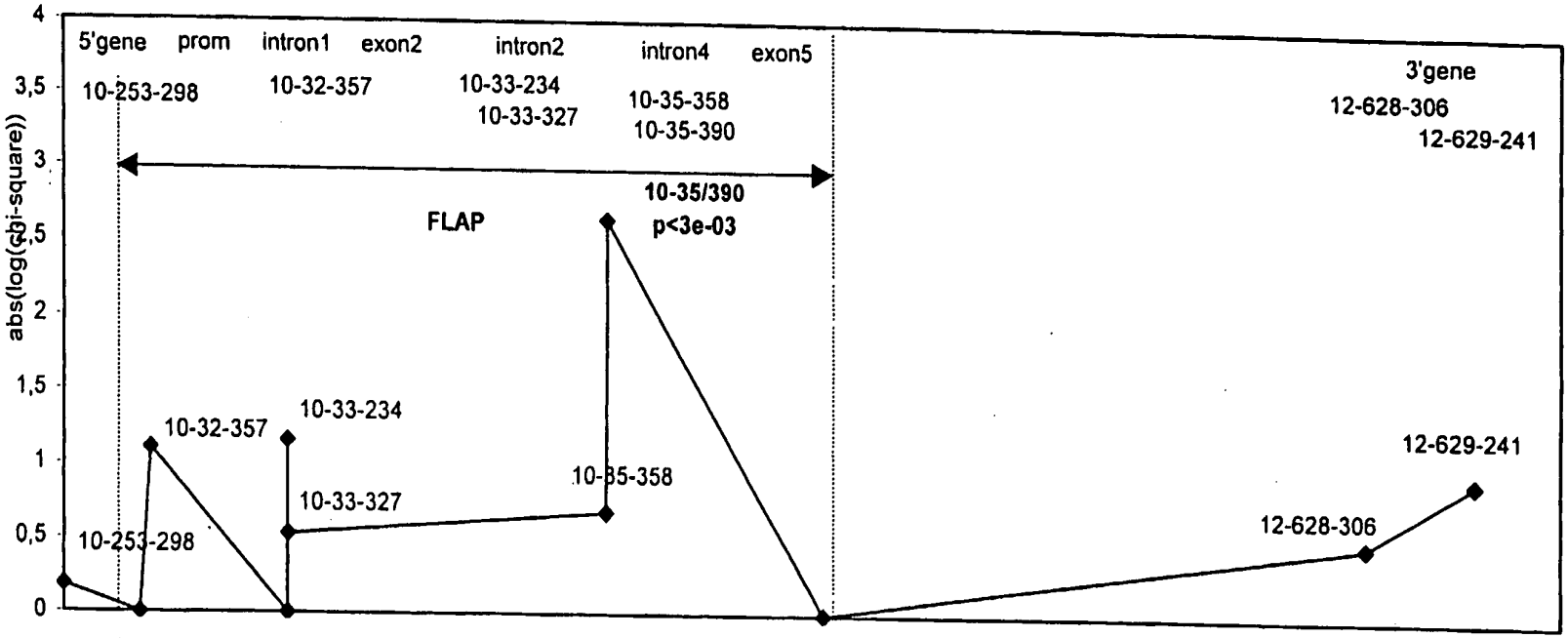
09

Printed: 29-12-2004

FIG.11

ASSOCIATION STUDIES BETWEEN THE FLAP BIALLELIC MARKERS AND ASTHMA

290 asthamatic individuals vs 280 caucasian US controls



DRAW

PCT / IB 0 0 / 0

IB000184

28-02-2000

61

FIG.12

HAPLOTYPE FREQUENCY ANALYSIS

MARKERS		10-253-298	10-33-175	10-33-234	10-33-327	10-35-358	10-35-390	12-628-306	12-629-241	ESTIMATED FREQUENCIES				
FLAP		5' gene	exon 2	intron 2		intron 4		3' gene		Frequencies haplotype				
cases/controls		287/186	295/174	295/274	295/270	291/280	295/272	284/185	283/182					
freq % case/controls		95/95 (C)	99/98 (C)	49/44 (A)	78/76 (T)	72/69 (G)	31/23 (C)	88/90 (C)	76/72 (G)					
diff.freq.all. (cases - controls)		0,5	1,8	5,3	2,6	3,4	9	2,1	4,6					
pvalue		6,55E-01	1,35E-02	6,93E-02	2,94E-01	2,06E-01	2,29E-03	3,17E-01	1,14E-01	cases	controls	Odds ratio	Chi-S	Pvalue (1df)
1	293 vs 265			A			T			0.283	0.197	1.61	11.18	(8.2e-04)
2	281 vs 177			A					G	0.305	0.210	1.65	9.97	(1.6e-03)
3	293 vs 261				T		T			0.307	0.224	1.53	9.62	(1.8e-03)
4	289 vs 271					G	T			0.304	0.231	1.46	7.77	(5.2e-03)
5	293 vs 168	C					T			0.309	0.226	1.53	7.26	(6.9e-03)
6	293 vs 265			A	T					0.276	0.208	1.46	7.17	(7.3e-03)
7	282 vs 178						T		G	0.314	0.233	1.50	7.01	(7.7e-03)
37	281 vs 176			A			T	C		0.265	0.171	1.76	11.04	(8.6e-04)
38	280 vs 173			A			T		G	0.292	0.194	1.71	10.71	(1.0e-03)
39	289 vs 264			A		G	T			0.283	0.199	1.59	10.56	(1.1e-03)
40	278 vs 175			A				C	G	0.271	0.180	1.70	9.94	(1.6e-03)
41	284 vs 176	C		A			T			0.287	0.195	1.66	9.77	(1.7e-03)
121	277 vs 171			A			T	C	G	0.265	0.169	1.77	11.07	(8.6e-04)
122	278 vs 173			A		G	T		G	0.290	0.195	1.69	10.29	(1.3e-03)
123	279 vs 176			A		G	T	C		0.264	0.175	1.70	9.80	(1.7e-03)
124	276 vs 175			A		G		C	G	0.271	0.181	1.69	9.72	(1.7e-03)
125	280 vs 174	C		A			T	C		0.265	0.176	1.69	9.68	(1.8e-03)
247	275 vs 171			A		G	T	C	G	0.265	0.170	1.77	10.91	(9.1e-04)
248	276 vs 169	C		A			T	C	G	0.265	0.172	1.74	10.30	(1.3e-03)
373	274 vs 169	C		A		G	T	C	G	0.265	0.172	1.73	10.13	(1.4e-03)
457	273 vs 163	C		A	T	G	T	C	G	0.247	0.167	1.64	7.74	(5.2e-03)

28-02-2000

Printed:29-12-2004

DRAW

PCT / IB 0 0 / 0 IB0000184

62

FIG.13

HAPLOTYPE FREQUENCY ANALYSIS
PERMUTATIONS TEST RESULTS (>1000 Iterations)

Markers	10-33-234 intron 2		10-35-390 intron 4	
ALT vs US	A		T	
cases vs US controls	5.3 (51 vs 56)	6,93E-02	9 (31 vs 23)	2,29E-03
ASSOCIATION	diff all. Freq	pvalue	diff all. Freq	pvalue

HAPLOTYPE (AT)	sample sizes cases vs controls	haplotype frequencies		p-excess	odds-ratio	chi-S	P value	PERMUTATIONS TEST RESULTS		
		cases	controls					Av. Chi-S	Max Chi-S	> Iter / nb of Iter.
Asthmatics vs US controls								1,2	7,4	0/1000
	293 vs 265	0,283	0,197	10,7	1,61	11,18	8,20E-04	1,2	12,9	1/10 000

Printed: 29-12-2004
DRAW
62/76
PCT / IB 0 0 / 0 IB0000184
28-02-2000

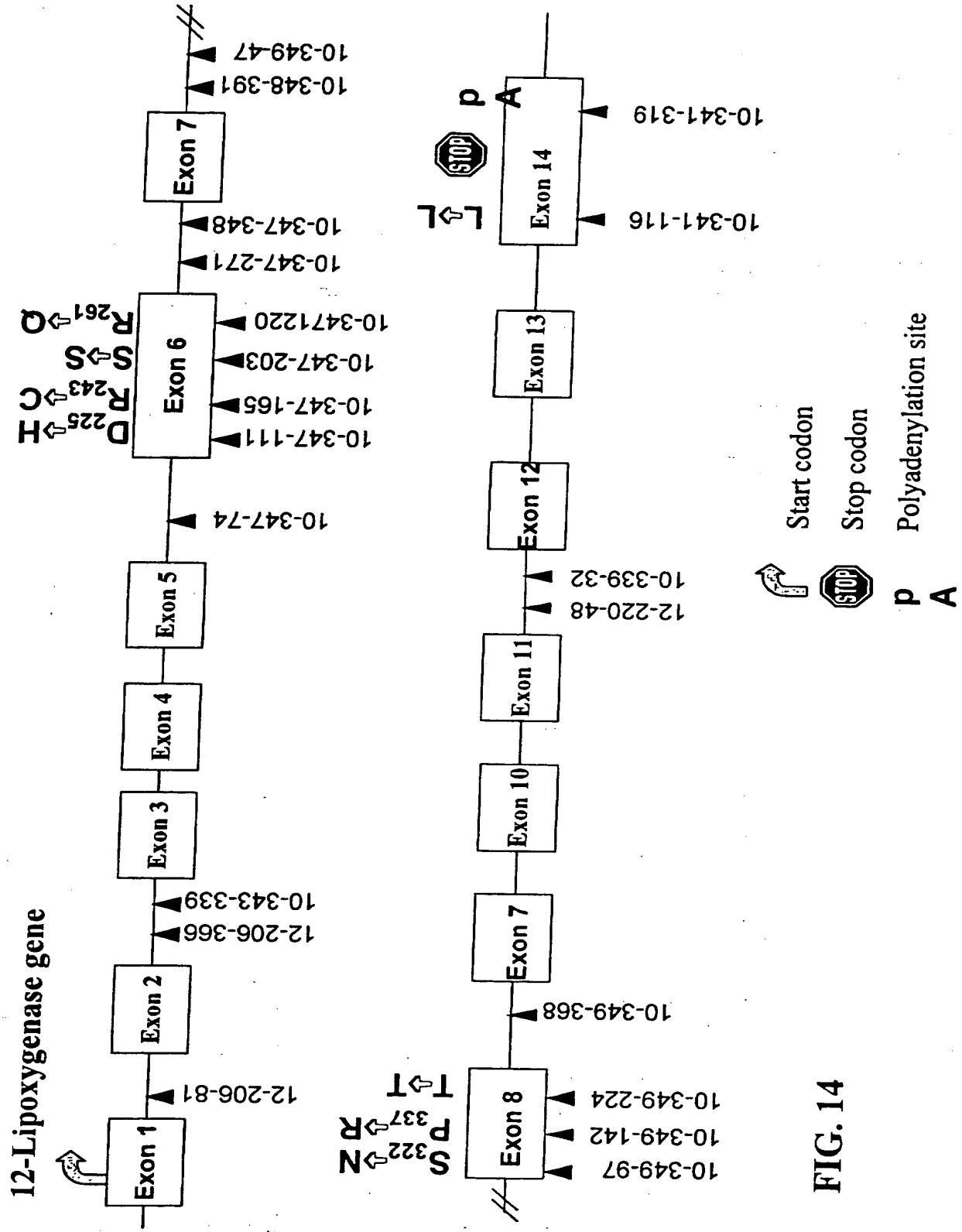

12-Lipoxygenase gene
Exon 1
12-206-81
Exon 2
12-206-366
10-343-339
Exon 3
Exon 4
Exon 5
10-347-74
Exon 6
10-347-111
10-347-165
10-347-203
10-34712 20
10-347-271
10-347-348
Exon 7
10-348-391
10-349-47
Exon 8
10-349-97
10-349-142
10-349-224
10-349-368
Exon 7
Exon 10
Exon 11
12-220-48
10-339-32
Exon 12
Exon 13
10-341-116
Exon 14
10-341-319
Start codon
Stop codon
Polyadenylation site

FIG. 14

FIG.15

HAPLOTYPE FREQUENCY ANALYSIS
(Asthma)

297 Asthmatics vs 186 US controls randoms

MARKERS	12-208-35	12-226-167	12-206-366	10-347-203	10-347-220	10-349-97	10-349-224	12-196-119	12-214-129	12-216-421	12-219-230	12-223-207	ESTIMATED FREQUENCIES					
12-lipoxygenase	5' gene		intron 2	exon 6		exon 8												
cases / controls	284/182	288/188	272/89	285/184	274/184	282/182	271/177	281/184	282/181	288/182	288/187	287/186	haplotype frequencies					
frequency % (case/controls)	59/58 (T)	62/59 (C)	57/62 (T)	57/58 (A)	58/60 (G)	59/60 (A)	57/60 (G)	70/71 (T)	61/61 (T)	61/64 (G)	64/68 (A)	62/62 (T)						
diff freq. all. (cases controls)	0.9	3 4	-4 6	-1 2	-1 7	-1 9	-3.1	-1.2	-0.7	-2.9	-4.2	0 8						
pvalue	7 52e-01 *	2 94e-01 *	2.73e-01 *	6.55e-01 *	5.84e-01 *	5 27e-01 *	3.43e-01 *	6 55e-01 *	7 52e-01 *	3 71e-01 *	1.80e-01	7 52e-01 *	cases	controls	p-excess	Odds ratio	Chi-S	Pvalue (1df)
1 268 vs 176					G				C		G		0.123	0.040	8.63	3.38	17.85	(2.3e-05)
2 277 vs 174						A			C		G		0.125	0 041	8.71	3.31	17.75	(2.5e-05)
3 274 vs 179					G					A	G		0.123	0 041	8.49	3 26	17.47	(2 9e-05)
4 282 vs 176						A				A	G		0 125	0.043	8.57	3 20	17 29	(3.2e-05)
5 280 vs 176				A					C		G		0 115	0.037	8.08	3 36	16 81	(3 9e-05)
6 285 vs 178				A						A	G		0.113	0 039	7 73	3.16	15.62	(7 4e-05)
7 270 vs 176		C					T					T	0.130	0.055	7.98	2 58	13.40	(2.5e-04)
8 247 vs 86			C		A		T						0 405	0.256	19.96	1 97	12 10	(5 0e-04)
9 255 vs 85			C	G			T						0 406	0.259	19.81	1.95	11.80	(5 6e-04)
10 253 vs 84			C			G	T						0.399	0 253	19 62	1 97	11.73	(5.9e-04)
11 267 vs 172	A						T					T	0.088	0 030	5.97	3 09	11.45	(7 0e-04)
12 281 vs 181		C				G						T	0 136	0 066	7.50	2.22	11 10	(8 6e-04)
13 274 vs 182		C			A							T	0.137	0 067	7.52	2.21	11.05	(8.6e-04)
14 278 vs 174				A				T		A	G		0.118	0.031	9.01	4.18	21.01	(4.4e-06)
15 267 vs 175					G			T		A	G		0 124	0.035	9 27	3 92	20.87	(4 8e-06)
16 276 vs 173						A		T		A	G		0 124	0.035	9.23	3 91	20 65	(5 4e-06)
17 273 vs 172				A				T	C		G		0.121	0.034	9.01	3.90	20.02	(7 3e-06)
18 268 vs 172					G				C	A	G		0.124	0 036	9.14	3.76	19.84	(8 2e-06)
19 261 vs 172					G			T	C		G		0.126	0.037	9.20	3.74	19.81	(8.2e-06)
20 271 vs 171						A		T	C		G		0.125	0 037	9.11	3 69	19 49	(1 0e-05)
21 277 vs 169						A			C	A	G		0.125	0.038	9.06	3.64	19.10	(1 2e-05)
22 280 vs 171				A					C	A	G		0 116	0.033	8.56	3 81	18.76	(1.5e-05)
23 264 vs 170					G.	A			C		G		0.125	0.040	8.91	3.45	18.15	(2.0e-05)
24 264 vs 81	T		C			G				G			0.197	0.056	14.96	4.13	18.01	(2.1e-05)

Printed:29-12-2004
PRAVI
PCT / IB 0 0 / 0
IB0000184
28-02-2000

FIG.16A

ALLELE FREQUENCY ANALYSIS
(Asthma)

CASES (297 ALT) vs CONTROLS (186 US CAUCASIAN)

MARKERS		12-197/244	12-208/35	12-226/167	12-206/366	10-346/141	10-347/111	10-347/165	10-347/203	10-347/220
PROTEIN 12-LO		5'gene			in2	ex5	ex6			
cases / controls		277/180	284/182	288/188	272/89	285/185	284/180	268/185	280/184	283/184
frequency % (case/controls)		66/67 (T)	58/57 (T)	62/58 (C)	57/61 (T)	99/100 (G)	99/100 (G)	99/100 (C)	57/58 (A)	57/59 (G)
diff freq. all. (cases - controls)		-1.0	0.9	3.4	-4.6	-0.4	-0.2	-0.2	-1.1	-2.1
pvalue		7.52e-01 *	7.52e-01 *	2.94e-01 *	2.73e-01 *	HOM	HOM	5.92e-01# *	6.55e-01 *	4.80e-01 *
Test Hardy Weinberg	cases vs	0.034 (HWD)	-0.002 (HWE)	-0.001 (HWE)	-0.014 (HWE)	0.000 (HWD)	0.000 (HWD)	0.000 (HWD)	-0.011 (HWE)	-0.005 (HWE)
	controls	0.054 (HWD)	-0.020 (HWE)	0.022 (HWE)	0.000 (HWE)	0.000 (HWD)	0.000 (HWD)	0.000 (HWD)	0.012 (HWE)	0.021 (HWE)

MARKERS		10-349/97	10-349/224	10-341/116	12-196/119	12-214/129	12-216/421	12-219/230	12-223/207
PROTEIN 12-LO		ex8		ex14	*markers in bac*				
cases / controls		287/182	277/177	286/176	281/184	282/181	288/182	288/187	287/186
frequency % (case/controls)		59/60 (A)	56/60 (G)	89/89 (G)	69/70 (T)	60/61 (T)	61/64 (G)	63/67 (A)	62/61 (T)
diff freq. all. (cases - controls)		-1.4	-4.1	0.1	-1.2	-0.7	-2.9	-4.2	0.8
pvalue		6.55e-01 *	2.06e-01 *	7.52e-01 *	6.55e-01 *	7.52e-01 *	3.71e-01 *	1.80e-01 *	7.52e-01 *
Test Hardy Weinberg	cases vs	0.003 (HWE)	-0.010 (HWE)	-0.008 (HWE)	0.012 (HWE)	-0.013 (HWE)	-0.012 (HWE)	-0.010 (HWE)	0.012 (HWD)
	controls	0.008 (HWE)	-0.004 (HWE)	-0.000 (HWE)	0.030 (HWE)	0.016 (HWE)	0.024 (HWE)	-0.001 (HWE)	-0.019 (HWD)

Printed: 29-12-2004
DRAWN
PCT / IB 0 0 / 0 0 0 1 8 4
IB0000184
28-02-2000

66

FIG.16B

HAPLOTYPE FREQUENCY ANALYSIS (Asthma)

CASES (297 ALT) vs CONTROLS (186 US CAUCASIAN)

			Marker 1	Marker 2	Marker 3	Marker 4	Haplotype	ESTIMATED FREQUENCIES: Haplotype frequencies cases	controls	p-excess	Odds ratio	Chi-S	Pvalue (1df)		PERMUTATIONS TEST RESULTS: Av. Chi-S	Max Chi-s	> Iter / nb of Iter
haplotype 1	PT2	265 vs 86	12-206/366	10-349/224			CT	0.424	0.265	21.72	2.05	13.97	(1.8e-04)	****	2.2	8 3	[0/100]
haplotype 2	PT2	267 vs 89	12-206/366	10-347/220			CA	0.423	0.274	20 56	1.94	12.55	(3.9e-04)	***	2.3	9.3	[0/100]
haplotype 3	PT2	266 vs 88	12-206/366	10-347/203			CG	0 421	0 277	19.96	1.90	11 64	(6 3e-04)	***	2 1	7 1	[0/100]
haplotype 4	PT2	271 vs 87	12-206/366	10-349/97			CG	0.408	0.270	18.88	1.86	10.69	(1.1e-03)	***	1 7	5 9	[0/100]
haplotype 5	PT2	271 vs 174	12-197/244	12-214/129			CC	0.148	0.077	7.66	2.08	10 03	(1.5e-03)	***	1.8	9.7	[0/100]
haplotype 6	PT2	285 vs 175	10-341/116	12-223/207			AT	0.042	0.008	3.45	5 48	8.95	(2.7e-03)	***	1.4	9.9	[2/100]
haplotype 7	PT3	282 vs 174	10-349/97	12-214/129	12-219/230		ACG	0.125	0 041	8.73	3.32	17 87	(2.3e-05)	*****	1.8	8 5	[0/100]
haplotype 8	PT3	287 vs 176	10-349/97	12-216/421	12-219/230		AAG	0 126	0.043	8.67	3.23	17.65	(2.6e-05)	*****	1.5	13.6	[0/100]
haplotype 9	PT3	277 vs 176	10-347/220	12-214/129	12-219/230		GCG	0.119	0.040	8.28	3.27	16.90	(3 7e-05)	****	3.0	15 4	[0/100]
haplotype 10	PT3	275 vs 176	10-347/203	12-214/129	12-219/230		ACG	0.115	0 037	8.05	3 35	16.64	(4.4e-05)	****	2.4	19.2	[1/100]
haplotype 11	PT3	283 vs 179	10-347/220	12-216/421	12-219/230		GAG	0 119	0 041	8 11	3.15	16.40	(5.1e-05)	****	2.6	14.4	[0/100]
haplotype 12	PT3	266 vs 171	12-197/244	10-347/203	12-214/129		CAC	0 070	0.012	5.96	6.46	15.97	(6.3e-05)	****	1.9	11.3	[0/100]
haplotype 13	PT3	248 vs 85	12-206/366	10-347/165	10-349/224		CCT	0.427	0.255	23.11	2 18	15.86	(6.7e-05)	****	2.7	18.8	[1/100]
haplotype 14	PT3	271 vs 168	12-197/244	10-349/97	12-214/129		CAC	0.069	0.012	5.83	6.37	15.34	(8.7e-05)	****	1 8	10.4	[0/100]
haplotype 15	PT3	261 vs 86	12-206/366	10-347/220	10-349/224		CAT	0.423	0 256	22.45	2.13	15.27	(9.2e-05)	****	2.4	7.8	[0/100]
haplotype 16	PT3	276 vs 176	12-226/167	10-349/224	12-223/207		CTT	0 137	0.055	8.64	2.72	15.27	(9.2e-05)	****	1 4	7.1	[0/100]
haplotype 17	PT3	280 vs 178	10-347/203	12-216/421	12-219/230		AAG	0 112	0.039	7.59	3.12	15.15	(9 7e-05)	****	2.2	21.1	[2/100]
haplotype 18	PT3	268 vs 170	12-197/244	10-347/220	12-214/129		CGC	0.067	0.012	5.60	6 10	14 68	(1.3e-04)	****	2.0	12.7	[0/100]
haplotype 19	PT3	249 vs 88	12-206/366	10-347/165	10-347/220		CCA	0 428	0.265	22 14	2.07	14.54	(1.3e-04)	****	2 8	9.4	[0/100]
haplotype 20	PT3	264 vs 86	12-206/366	10-346/141	10-349/224		CGT	0 426	0 265	21.94	2.06	14.23	(1 6e-04)	****	2 6	11.2	[0/100]
haplotype 21	PT3	261 vs 85	12-206/366	10-347/203	10-349/224		CGT	0.418	0 259	21 40	2.05	13.71	(2 0e-04)	****	2.2	8 3	[0/100]
haplotype 22	PT3	264 vs 84	12-206/366	10-349/97	10-349/224		CGT	0 411	0.253	21.19	2 06	13 68	(2.1e-04)	****	2.4	7.2	[0/100]
haplotype 23	PT3	248 vs 87	12-206/366	10-347/165	10-347/203		CCG	0 425	0 268	21.54	2.03	13.55	(2.3e-04)	****	2 2	8.5	[0/100]
haplotype 24	PT3	261 vs 86	12-206/366	10-347/111	10-349/224		CGT	0.421	0 265	21.30	2.02	13 43	(2.4e-04)	****	2.1	6 1	[0/100]
haplotype 25	PT3	268 vs 164	12-197/244	10-347/111	12-214/129		CGC	0 151	0.068	8.89	2.44	13.33	(2 5e-04)	****	1 9	25.3	[1/100]
haplotype 26	PT3	265 vs 89	12-206/366	10-346/141	10-347/220		CGA	0.426	0 274	21.00	1.97	13.04	(3.0e-04)	****	2.3	7.0	[0/100]
haplotype 27	PT4	280 vs 173	10-349/97	12-196/119	12-216/421	12-219/230	ATAG	0.124	0.035	9.20	3.90	20.63	(5.4e-06)	*****	1.5	11 6	[0/100]
haplotype 28	PT4	274 vs 174	10-347/203	12-196/119	12-216/421	12-219/230	ATAG	0.117	0.031	8.89	4.14	20.59	(5.7e-06)	*****	2.7	19.0	[0/100]
haplotype 29	PT4	275 vs 171	10-349/97	12-196/119	12-214/129	12-219/230	ATCG	0.126	0 037	9.21	3.72	19.86	(8.2e-06)	*****	2 0	11.8	[0/100]
haplotype 30	PT4	276 vs 175	10-347/220	12-196/119	12-216/421	12-219/230	GTAG	0.121	0.035	8.92	3.80	19.84	(8.2e-06)	*****	2.4	14.7	[0/100]
haplotype 31	PT4	269 vs 172	10-347/203	12-196/119	12-214/129	12-219/230	ATCG	0.120	0.034	8 90	3.86	19.61	(9.1e-06)	*****	2.4	12 2	[0/100]
haplotype 32	PT4	280 vs 165	10-349/97	10-341/116	12-214/129	12-219/230	AGCG	0.127	0.038	9.30	3 73	19 55	(9.5e-06)	*****	1.9	16.0	[0/100]
haplotype 33	PT4	270 vs 172	10-347/220	12-196/119	12-214/129	12-219/230	GTCG	0 124	0.037	9.00	3.68	19.31	(1 1e-05)	*****	2 8	19.3	[0/100]
haplotype 34	PT4	282 vs 169	10-349/97	12-214/129	12-216/421	12-219/230	ACAG	0.124	0.038	9.00	3.62	18.98	(1.3e-05)	*****	1.8	10.5	[0/100]
haplotype 35	PT4	267 vs 167	12-197/244	12-208/35	12-214/129	12-223/207	CTCC	0 055	0.000	5.49	100.00	18 96	(1.3e-05)	*****	2 8	21.7	[2/100]
haplotype 36	PT4	285 vs 167	10-349/97	10-341/116	12-216/421	12-219/230	AGAG	0 127	0.039	9.12	3.55	18.89	(1 4e-05)	*****	2.0	12.3	[0/100]
haplotype 37	PT4	277 vs 172	10-347/220	12-214/129	12-216/421	12-219/230	GCAG	0.120	0.036	8.71	3.62	18.59	(1.6e-05)	*****	2.3	25.8	[1/100]
haplotype 38	PT4	275 vs 171	10-347/203	12-214/129	12-216/421	12-219/230	ACAG	0.116	0.033	8.52	3.80	18.57	(1.6e-05)	*****	2.7	21.6	[1/100]
haplotype 39	PT4	276 vs 162	12-208/35	10-341/116	12-214/129	12-219/230	AGCG	0.054	0.000	5.40	0 00	18.10	(2.0e-05)	*****	2.7	18.6	[1/100]
haplotype 40	PT4	245 vs 85	12-206/366	10-347/165	10-347/220	10-349/224	CCAT	0.429	0.246	24.18	2.30	17.77	(2 5e-05)	*****	3.0	7.5	[0/100]
haplotype 41	PT4	268 vs 81	12-208/35	12-206/366	10-349/97	12-216/421	TCGG	0 196	0 056	14.80	4.09	17.78	(2.5e-05)	*****	1.8	15.1	[0/100]

Printed:29-12-2004
DRAWI
66/76
PCT / IB 0 0 / IB0000184
28-02-2000
IB0000184

FIG.17

HAPLOTYPE FREQUENCY ANALYSIS
(Zyflo secondary effects)

89 ALT+ vs 208 ALT-

MARKERS	12-208-35	12-226-167	12-206-366	10-347-203	10-347-220	10-349-97	10-349-224	12-196-119	12-214-129	12-216-421	12-219230	12-223-207	ESTIMATED FREQUENCIES				
12-lipoxygenase	5' gene		intron 2	exon 6		exon 8											
Size (cases / controls)	87/197	89/199	86/186	88/197	86/188	86/196	86/185	86/195	89/193	89/199	89/199	88/199	haplotype frequencies				
frequency % (case/controls)	58/59 (T)	61/63 (C)	55/58 (T)	56/58 (A)	56/59 (G)	58/59 (A)	54/59 (G)	72/69 (T)	59/61 (T)	58/63 (G)	67/62 (A)	62/63 (T)					
diff freq. all. (cases - controls)	-0.8	-2 1	-3 7	-2 2	-3 0	-1 4	-4 8	3 4	-2.4	-4.7	4 5	-0.6					
p value	7 52e-01 *	5.84e-01 *	4.03e-01 *	5 84e-01 *	4 80e-01 *	7 52e-01 *	2 73e-01 *	4 03e-01 *	5 84e-01 *	2 73e-01 *	2.94e-01	7 52e-01 *	cases	controls	Odds ratio	Chi-S	Pvalue (1df)
1 87 vs 197	A	G											0 123	0 070	1.87	4.34	(3.6e-02)
2 83 vs 184				A				C					0 209	0 140	1.63	4 09	(4.3e-02)
3 85 vs 185							T			A		T	**0.151**	**0.048**	**3.53**	**16.76**	**(4.2e-05)**
4 85 vs 186			C							A		T	0.148	0.059	2.77	11.62	(6.3e-04)
5 85 vs 179							T		C			T	0 151	0.066	2.52	9 85	(1 7e-03)
6 85 vs 180			C						C			T	0 149	0.070	2.34	8.42	(3.6e-03)
7 86 vs 188				A						A		T	0.157	0.076	2.25	8.36	(3.8e-03)
8 82 vs 174			C				T			A		T	**0.158**	**0.040**	**4.56**	**21.85**	**(2.9e-06)**
9 85 vs 179							T		C	A		T	0 157	0.052	3 37	16 03	(6.0e-05)
10 83 vs 177				A			T			A		T	0.162	0 057	3.22	15 30	(9.2e-05)
11 82 vs 183						G	T			A		T	0.147	0.050	3 23	14.18	(1 6e-04)
12 85 vs 183				G			T			A		T	0.143	0.049	3.21	13 97	(1.8e-04)
13 82 vs 168			C				T		C			T	0.156	0.058	3 02	13.08	(3.0e-04)
14 84 vs 175			C		A					A		T	0.155	0.062	2.78	11.77	(5.9e-04)
15 84 vs 184			C	G						A		T	0 147	0 059	2.74	11.22	(7.8e-04)
16 85 vs 180			C						C	A		T	0.145	0.058	2 74	11 05	(8.6e-04)
17 82 vs 181							T	T		A		T	0 132	0.050	2 85	10 62	(1.1e-03)
18 83 vs 187	A							T	T			T	0.148	0.066	2.48	9.50	(2.1e-03)
19 83 vs 171				A			T		C			T	0.162	0.074	2.41	9.29	(2.3e-03)
20 83 vs 174			C				T			A	A		0.111	0.041	2.92	9.24	(2.3e-03)
21 82 vs 178						G	T		C			T	0 152	0 068	2.45	9 22	(2.3e-03)
22 82 vs 168			C				T		C	A		T	0.161	0 043	4.27	20.43	(6.0e-06)
23 82 vs 172			C	G			T			A		T	0.146	0.040	4.07	18.03	(2 1e-05)
24 81 vs 166			C		A		T			A		T	0 160	0 047	3.82	17.77	(2 5e-05)
25 80 vs 171			C				T	T		A		T	0 137	0.037	4.17	17.18	(3 4e-05)

Printed: 29-12-2004
DRAWN
PCT / IB 0 0 / 0
IB0000184
28022000

68

FIG.18A

ALLELE FREQUENCY ANALYSIS (Zyflo secondary effects)

CASES (85 ALT+) vs CONTROLS (208 ALT-)

MARKERS		12-197/244	12-208/35	12-226/167	12-206/366	10-346/141	10-347/111	10-347/165	10-347/203	10-347/220
PROTEIN 12-LO		5'gene			In2	ex5	ex6			
cases / controls		81/196	87/197	89/199	86/186	88/197	88/196	69/199	83/197	87/196
frequency % (case/controls)		70/65(T)	58/59(T)	61/63(C)	55/58(T)	100/99(G)	99/100(G)	100/99(C)	56/58(A)	43/57(G)
diff freq. all. (cases - controls)		5.8	-0.8	-2.1	-3.7	0.5	-0.6	0.3	-1.8	0.1
pvalue		1.80e-01 *	7.52e-01 *	5.84e-01 *	4.03e-01 *	HOM	HOM	7.43e-01# *	6.55e-01 *	7.52e-01 *
Test Hardy Weinberg	cases vs	-0.001 (HWE)	0.008 (HWE)	-0.020 (HWE)	-0.031 (HWE)	**0.000 (HWD)**	**0.000 (HWD)**	**0.000 (HWD)**	-0.037 (HWE)	-0.020 (HWE)
	controls	**0.048 (HWD)**	-0.007 (HWE)	0.007 (HWE)	-0.007 (HWE)	**0.000 (HWD)**	**0.000 (HWD)**	**0.000 (HWD)**	0.000 (HWE)	0.002 (HWE)

MARKERS		10-349/97	10-349/224	10-341/116	12-196/119	12-214/129	12-216/421	12-219/230	12-223/207
PROTEIN 12-LO		ex8		ex14	*markers in bac*				
cases / controls		89/198	83/194	89/197	86/195	89/193	89/199	89/199	88/199
frequency % (case/controls)		59/59(A)	54/57(G)	90/89(G)	72/69(T)	59/61(T)	58/63(G)	67/62(G)	62/63(T)
diff freq. all. (cases - controls)		-0.1	-3.0	1.6	3.4	-2.4	-4.7	4.5	-0.6
pvalue		7.52e-01 *	4.80e-01 *	5.27e-01 *	4.03e-01 *	5.84e-01 *	2.73e-01 *	2.94e-01 *	7.52e-01 *
Test Hardy Weinberg	cases vs	0.000 (HWE)	-0.029 (HWE)	0.002 (HWE)	0.015 (HWE)	-0.011 (HWE)	-0.031 (HWE)	0.002 (HWE)	0.037 (HWE)
	controls	0.004 (HWE)	-0.003 (HWE)	**0.764 (HWD)**	0.010 (HWE)	-0.014 (HWE)	-0.004 (HWE)	-0.016 (HWE)	0.001 (HWE)

Printed:29-12-2004
DRAW:
68/76
PCT / IB 0 0 / 0 IB000184
28-02-2000

FIG.18B

HAPLOTYPE FREQUENCY ANALYSIS
(Zyflo secondary effects)

CASES (85 ALT+) vs CONTROLS (208 ALT-)

			MARKER 1	MARKER 2	MARKER 3	MARKER 4	MARKER 5	HAPLOTYPE	ESTIMATED FREQUENCIES							PERMUTATIONS TEST RESULTS		
									Haplotype frequencies		p-excess	Odds ratio	Chi-S	Pvalue (1df)		Av. Chi-S	Max Chi-s	> Iter / No. of Iter
									cases	controls								
haplotype 1	PT2	79 vs 192	12-197/244	12-196/119				TT	0.542	0 436	18 89	1.53	5 11	(2.3e-02)	**	1.7	6 7	[3/100]
haplotype 2	PT2	87 vs 197	12-208/35	12-226/167				AG	0.123	0.070	5.73	1.87	4.34	(3.6e-02)	*	1 1	7.3	[4/100]
haplotype 3	PT2	84 vs 183	12-206/366	12-196/119				CC	0.205	0.139	7.68	1.60	3 75	(5.1e-02)	*	1.3	8 0	[11/100]
haplotype 4	PT2	84 vs 192	10-347/220	12-196/119				GT	0 486	0 400	14 25	1.41	3 49	(6.1e-02)	*	1.0	5.3	[8/100]
haplotype 5	PT2	81 vs 193	10-347/203	12-196/119				GC	0.207	0 144	7.46	1.56	3.43	(6.1e-02)	*	1.3	10.3	[9/100]
haplotype 6	PT3	82 vs 194	10-349/224	12-216/421	12-223/207			TAT	0 158	0.064	10.09	2.76	12.35	(4.3e-04)	***	2.1	11.0	[0/100]
haplotype 7	PT3	85 vs 186	12-206/366	12-216/421	12-223/207			CAT	0.148	0.059	9.44	2.77	11.62	(6.3e-04)	***	2.8	24 7	[3/100]
haplotype 8	PT3	77 vs 180	12-197/244	12-206/366	12-196/119			TTT	0.434	0 286	20 67	1.91	10.62	(1.1e-03)	***	1.8	10.9	[1/100]
haplotype 9	PT3	78 vs 190	12-197/244	10-347/220	12-196/119			TGT	0.433	0.291	19.96	1.86	9 98	(1.6e-03)	***	1.7	9 3	[0/100]
haplotype 10	PT3	76 vs 187	12-197/244	10-349/224	12-196/119			TGT	0 435	0.293	20.17	1.86	9.88	(1.7e-03)	***	1.8	6.9	[0/100]
haplotype 11	PT3	77 vs 191	12-197/244	10-349/224	12-216/421			CTA	0.137	0 056	8.56	2.66	9.76	(1 7e-03)	***	1.6	11.3	[1/100]
haplotype 12	PT3	75 vs 191	12-197/244	10-347/203	12-196/119			TAT	0 431	0 294	19.42	1.82	9.13	(2 4e-03)	***	1.3	6 1	[0/100]
haplotype 13	PT4	81 vs 183	12-206/366	10-349/224	12-216/421	12-223/207		CTAT	0.160	0.058	10.82	3.10	14.38	(1.5e-04)	****	2.0	13 9	[0/100]
haplotype 14	PT4	84 vs 185	12-206/366	10-346/141	12-216/421	12-223/207		CGAT	0 158	0.058	10.62	3.04	14.20	(1.6e-04)	****	3.3	23.7	[2/100]
haplotype 15	PT4	82 vs 188	10-349/224	12-214/129	12-216/421	12-223/207		TCAT	0.161	0.063	10.50	2 86	13.12	(2.8e-04)	****	2.9	25.2	[3/100]
haplotype 16	PT4	81 vs 184	12-206/366	10-347/203	12-216/421	12-223/207		CGAT	0 153	0.059	10.03	2.89	12.50	(3.9e-04)	***	2.7	13.0	[1/100]
haplotype 17	PT4	82 vs 191	10-347/111	10-349/224	12-216/421	12-223/207		GTAT	0 159	0.064	10.14	2.77	12.33	(4.3e-04)	***	2.3	15.7	[2/100]
haplotype 18	PT4	82 vs 192	10-346/141	10-349/224	12-216/421	12-223/207		GTAT	0.158	0 065	9.96	2.70	11.85	(5.6e-04)	***	2.6	18 1	[3/100]
haplotype 19	PT4	81 vs 192	10-347/220	10-349/224	12-216/421	12-223/207		ATAT	0.159	0.066	10.02	2 69	11.78	(5.9e-04)	***	2.2	9.9	[0/100]
haplotype 20	PT4	84 vs 183	12-206/366	10-347/220	12-216/421	12-223/207		CAAT	0.150	0.059	9.60	2.79	11 72	(5.9e-04)	***	2.3	17.1	[3/100]
haplotype 21	PT4	85 vs 183	12-206/366	10-347/111	12-216/421	12-223/207		CGAT	0.148	0 059	9.42	2.76	11.45	(7.0e-04)	***	2.0	12.3	[1/100]
haplotype 22	PT4	76 vs 185	12-197/244	10-346/141	10-349/224	12-196/119		TGGT	0.435	0.284	21.12	1.94	11.16	(8.2e-04)	***	2.0	8.7	[0/100]
haplotype 23	PT4	85 vs 180	12-206/366	12-214/129	12-216/421	12-223/207		CCAT	0 145	0.058	9.20	2.74	11.03	(8.6e-04)	***	2.3	15.7	[3/100]
haplotype 24	PT4	80 vs 192	10-347/203	10-349/224	12-216/421	12-223/207		GTAT	0.156	0.065	9.65	2.63	11.00	(8 6e-04)	***	1.7	14.4	[3/100]
haplotype 25	PT4	82 vs 190	10-347/203	10-341/116	12-214/129	12-223/207		GGCT	0 125	0.046	8.22	2.94	10.86	(9.6e-04)	***	1.8	18.8	[2/100]
haplotype 26	PT5	77 vs 190	12-197/244	12-208/35	12-196/119	12-216/421	12-219/230	TATGA	0.138	0.050	9.27	3.06	12.24	(4.5e-04)	***	2.6	13.2	[2/100]
haplotype 27	PT5	77 vs 189	12-197/244	12-208/35	10-349/97	12-196/119	12-223/207	TTATC	0.127	0.045	8.56	3.08	11.42	(7 0e-04)	***	1.6	10.3	[0/100]
haplotype 28	PT5	77 vs 184	12-197/244	12-208/35	12-196/119	12-214/129	12-219/230	TATTA	0.126	0.047	8.30	2.93	10.42	(1.2e-03)	***	2.4	14.6	[4/100]
haplotype 29	PT5	76 vs 188	12-197/244	12-208/35	10-347/220	12-196/119	12-223/207	TTGTC	0.121	0.048	7.63	2.71	8.84	(2.9e-03)	***	1.4	9.5	[1/100]
haplotype 30	PT5	76 vs 176	12-197/244	12-208/35	12-206/366	10-341/116	12-196/119	TTTGT	0.195	0.099	10.67	2.21	8 82	(2.9e-03)	***	1 3	7.3	[0/100]

Printed: 29-12-2004
DRAWN
PCT / IB 0 0 / IB0000184
28-02-2000

FIG.19

Summary of
Association Study Results and Permutation Tests

12-Lipoxygenase	12-206-366	10-347-203	10-349-224	12-196-119	12-216-421	12-219-230	12-223-207	MARKERS	
	intron 2	exon 6	exon 8						
	C		T		A		T	HAPLOTYPE 8 Zyflo secondary effects (ALT+ vs ALT-)	
		A		T	A	G		HAPLOTYPE 14 Asthma (ALT vs US)	
	4,03E-01	5,84E-01	7,52E-01	4,03E-01	2,73E-01	2,94E-01	7,52E-01	pvalue	ALT+ vs ALT-
	-3,7	-2,2	-1,4	3,4	-4,7	4,5	-0,6	diff all. Freq	
	2,73E-01	6,55E-01	5,27E-01	6,55E-01	3,71E-01	1,80E-01	7,52E-01	pvalue	ALT vs caucasian US
	-4,6	-1,2	-1,9	-1,2	-2,9	-4,2	0,8	diff all. Freq	

HAPLOTYPE 8 (ALT+ vs ALT-) Zyflo secondary effects	sample sizes cases vs controls	haplotype frequencies		odds-ratio	chi-S	P value	PERMUTATIONS TEST RESULTS		
		cases	controls				Av. Chi-S	Max Chi-S	> Iter / nb of Iter.
ALT+ vs ALT-	82 vs 174	0,158	0,04	4,56	21,85	2,90E-06	3,1	29,9	5/1 000
							3,3	40,9	77/10 000
ALT vs caucasian US	256 vs 83	0,059	0		10,12	1,40E-03	3,5	37,6	82/1 000

HAPLOTYPE 14 (ALT vs US) Asthma	sample sizes cases vs controls	haplotype frequencies		odds-ratio	chi-S	P value	PERMUTATIONS TEST RESULTS		
		cases	controls				Av. Chi-S	Max Chi-S	> Iter / nb of Iter.
ALT+ vs ALT-	85 vs 193	0,097	0,109	-1,34	0,18	6,50E-01	2,1	24,1	785/1 000
ALT vs caucasian US							2,8	38,6	39/10 000
	278 vs 174	0,118	0,031	4,18	21,01	4,40E-06	2,8	29,9	7/1 000

Printed:29-12-2004
DRAW:
28-02-2000
PCT / IB 0 0 / IB0000184

FIG.20

PERMUTATIONS TEST RESULTS

PROTEIN PEC12L	12-206/366	10-349/97	10-349/224	12-196/119	12-214/129	12-216/421	12-219/230	12-223/207
	in2	ex8	ex8	in bac *(not localization in Bac: 3' or 5' gene)*				
	C		T			A		T
		A			C		G	
		A		T		A	G	
	4 03e-01	7 52e-01	4 80e-01	4 03e-01	5 84e-01	2 73e-01	2 94e-01	7.52e-01
	-3 7 (54 vs 58)	-0 1 (58 vs 59)	-3 0 (54 vs 57)	3 4 (72 vs 68)	-2 4 (58 vs 61)	-4 7 (57 vs 62)	4 5 (66 vs 62)	-0 6 (61 vs 62)
	2 73e-01	6 55e-01	2 06e-01	6.55e-01	7 52e-01	3 71e-01	1 80e-01	7 52e-01
	-4 6 (57 vs 61)	-1 4 (59 vs 60)	-4 1 (56 vs 60)	-1 2 (69 vs 70)	-0.7 (60 vs 61)	-2 9 (61 vs 64)	-4 2 (63 vs 67)	0 8 (62 vs 61)

MARKERS	
HAPLOTYPE 1	(ALT+ vs ALT-)
HAPLOTYPE 2 HAPLOTYPE 3	(ALT vs US)
pvalue diff all Freq (cases vs	ALT+ vs ALT-
pvalue diff all. Freq (cases vs	ALT vs caucasian US

Zyflo secondary effects

HAPLOTYPE 1 (ALT+ vs ALT-) (Zyflo secondary effects)CTAT	sample sizes cases vs controls	haplotype frequencies cases	haplotype frequencies controls	p-excess	odds-ratio	chi-S	P value		PERMUTATIONS TEST RESULTS Av. Chi-S	Max Chi-S	> Iter / nb of Iter.
ALT+ vs ALT-											
ALT+ vs ALT- (1)	81 vs 183	0,16	0,058	10,82	3,10	14,38	1,50E-04	****	2	13,9	0/100
ALT+ vs ALT- (2)	81 vs 99	0,16	0,065	10,11	2,72	8,28	4,00E-03	***	2,7	33,6	18/1000
ALT vs caucasian US	81 vs 84	0,16	0,044	12,12	4,15	12,23	4,50E-04	***	3,3	23,1	118/1000
	264 vs 83	0,071	0	302,77#	7,08	12,37	4,30E-04	***TH	2,7	19,6	20/1000

Asthma gene

HAPLOTYPE 2 (ALT vs US) (Asthma gene)ACG	sample sizes cases vs controls	haplotype frequencies cases	haplotype frequencies controls	p-excess	odds-ratio	chi-S	P value		PERMUTATIONS TEST RESULTS Av. Chi-S	Max Chi-S	> Iter / nb of Iter.
ALT+ vs ALT-											
ALT+ vs ALT- (1)	89 vs 193	0,131	0,121	1,14	1,10	0,11	6,50E-01	*	1,3	18,1	760/1000
ALT+ vs ALT- (2)	89 vs 104	0,131	0,115	1,84	1,16	0,20	5,80E-01	*	1,5	14,6	683/1000
ALT vs caucasian US	89 vs 89	0,131	0,134	-0,26	0,98	0,00	7,50E-01	*	1,4	16,2	946/1000
	282 vs 174	0,125	0,041	8,73	3,32	17,87	2,30E-05	*****	1,8	8,5	0/100
									2	19,9	2/1000

HAPLOTYPE 3 (ALT vs US) (Asthma gene) ATAG	sample sizes cases vs controls	haplotype frequencies cases	haplotype frequencies controls	p-excess	odds-ratio	chi-S	P value		PERMUTATIONS TEST RESULTS Av. Chi-S	Max Chi-S	> Iter / nb of Iter.
ALT+ vs ALT-											
ALT+ vs ALT- (1)	86 vs 194	0,123	0,114	1,04	1,09	0,10	7,50E-01	*	1,5	15,5	816/1000
ALT+ vs ALT- (2)	86 vs 100	0,123	0,108	1,69	1,16	0,21	5,80E-01	*	1,6	16,6	735/1000
ALT vs caucasian US	86 vs 94	0,123	0,11	1,54	1,14	0,16	6,50E-01	*	1,5	19,3	750/1000
	280 vs 173	0,124	0,035	9,2	3,9	20,63	5,40E-06	******	1,5	11,6	0/100
									2	18,7	0/1000

Printed: 29-12-2004
DRAW
71/76
PCT / IB 00 / IB0000184
28-02-2000

72

FIG.21

Allele Frequency

PROTEINS	Marker	ALT+					ALT-					US caucasian				
		size	A	C	G	T	size	A	C	G	T	size	A	C	G	T
12-LO	12-197/244	81		29.63		70.37	196		35.46		64.54	180		32.78		67.22
2	12-208/35	87	41.95			58.05	197	41.12			58.88	182	42.31			57.69
3	12-226/167	89		60.67	39.33		199		62.81	37.19		188		58.78	41.22	
4	12-206/366	86		45.35		54.65	186		41.67		58.33	89		38.20		61.80
5	10-346/141	88			HOM		197	0.51		99.49		185			HOM	
6	10-347/111	88		0.57	99.43		196			HOM		180			HOM	
7	10-347/165	69		HOM			199		99.75		0.25	185		HOM		
8	10-347/203	83	56.02		43.98		197	57.87		42.13		184	58.42		41.58	
9	10-347/220	87	42.53		57.47		196	42.60		57.40		184	40.49		59.51	
10	10-349/97	89	58.99		41.01		198	59.09		40.91		182	60.44		39.56	
11	10-349/224	83			54.22	45.78	194			57.22	42.78	177			60.45	39.55
12	10-341/116	89	9.55		90.45		197	11.17		88.83		176	10.80		89.20	
13	12-196/119	86		27.91		72.09	195		31.28		68.72	184		29.08		70.92
14	12-214/129	89		41.01		58.99	193		38.60		61.40	181		38.67		61.33
15	12-216/421	89	42.13		57.87		199	37.44		62.56		182	35.99		64.01	
16	12-219/230	89	66.85		33.15		199	62.31		37.69		187	67.91		32.09	
17	12-223/207	88		38.07		61.93	199		37.44		62.56	186		38.44		61.56

Printed: 29-12-2004
DRAW
72/76
PCT / IB 0 0 / 0
IB000184
28-02-2000

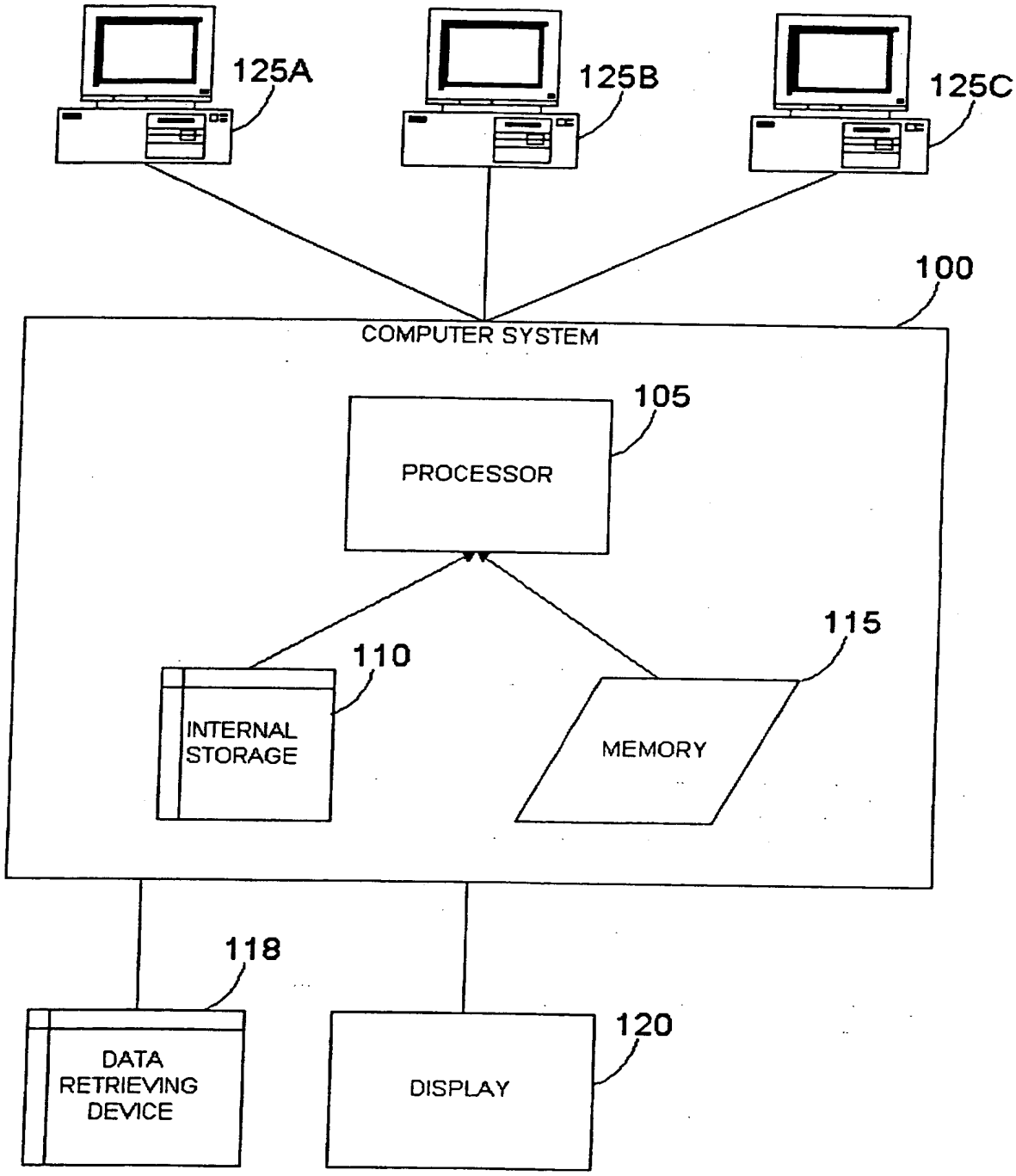

125A
125B
125C
100
COMPUTER SYSTEM
105
PROCESSOR
110
INTERNAL STORAGE
115
MEMORY
118
DATA RETRIEVING DEVICE
120
DISPLAY

FIG.22

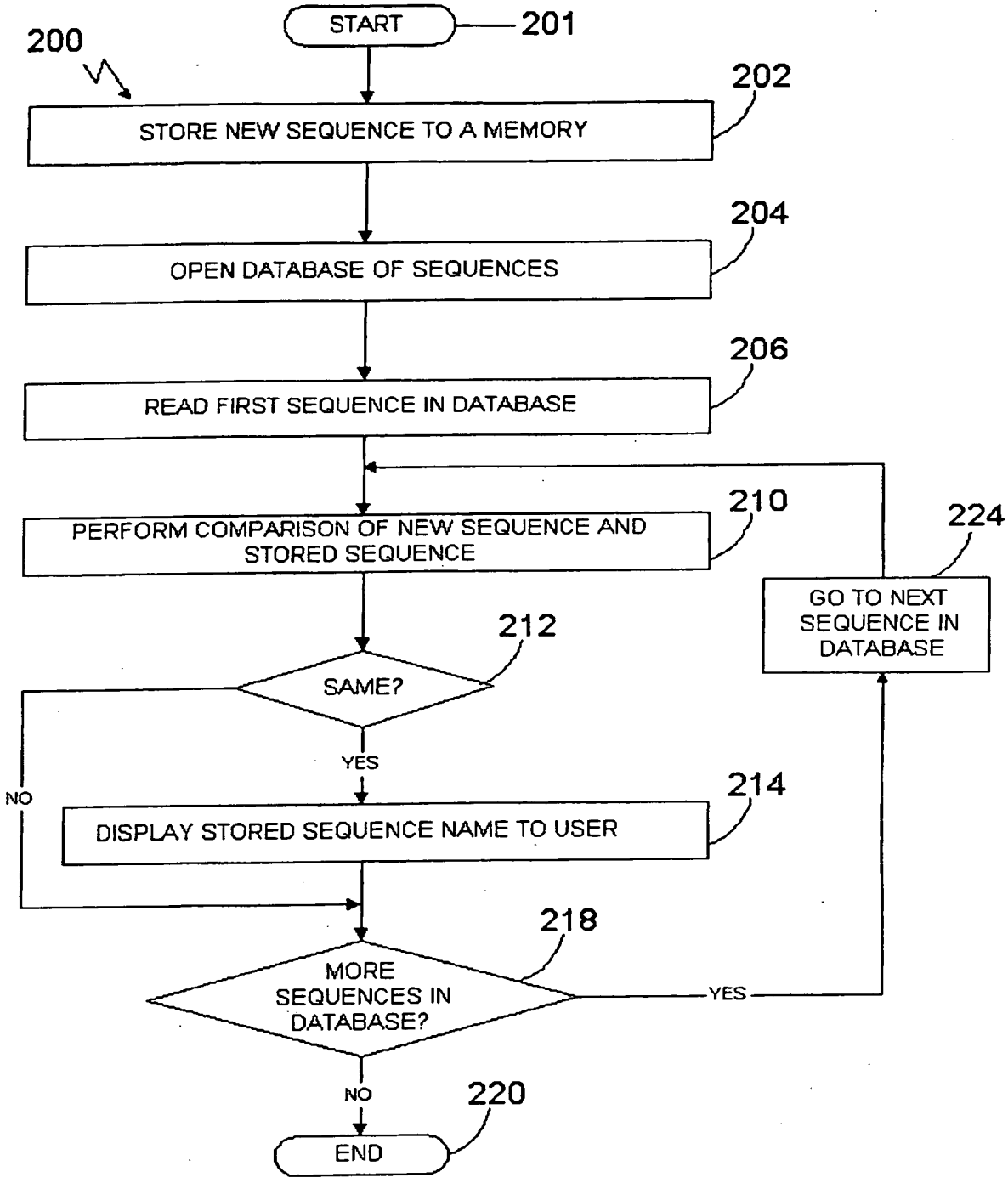

START
201
200
202
STORE NEW SEQUENCE TO A MEMORY
204
OPEN DATABASE OF SEQUENCES
206
READ FIRST SEQUENCE IN DATABASE
210
PERFORM COMPARISON OF NEW SEQUENCE AND STORED SEQUENCE
224
GO TO NEXT SEQUENCE IN DATABASE
212
SAME?
YES
NO
214
DISPLAY STORED SEQUENCE NAME TO USER
218
MORE SEQUENCES IN DATABASE?
YES
NO
220
END

FIG.23

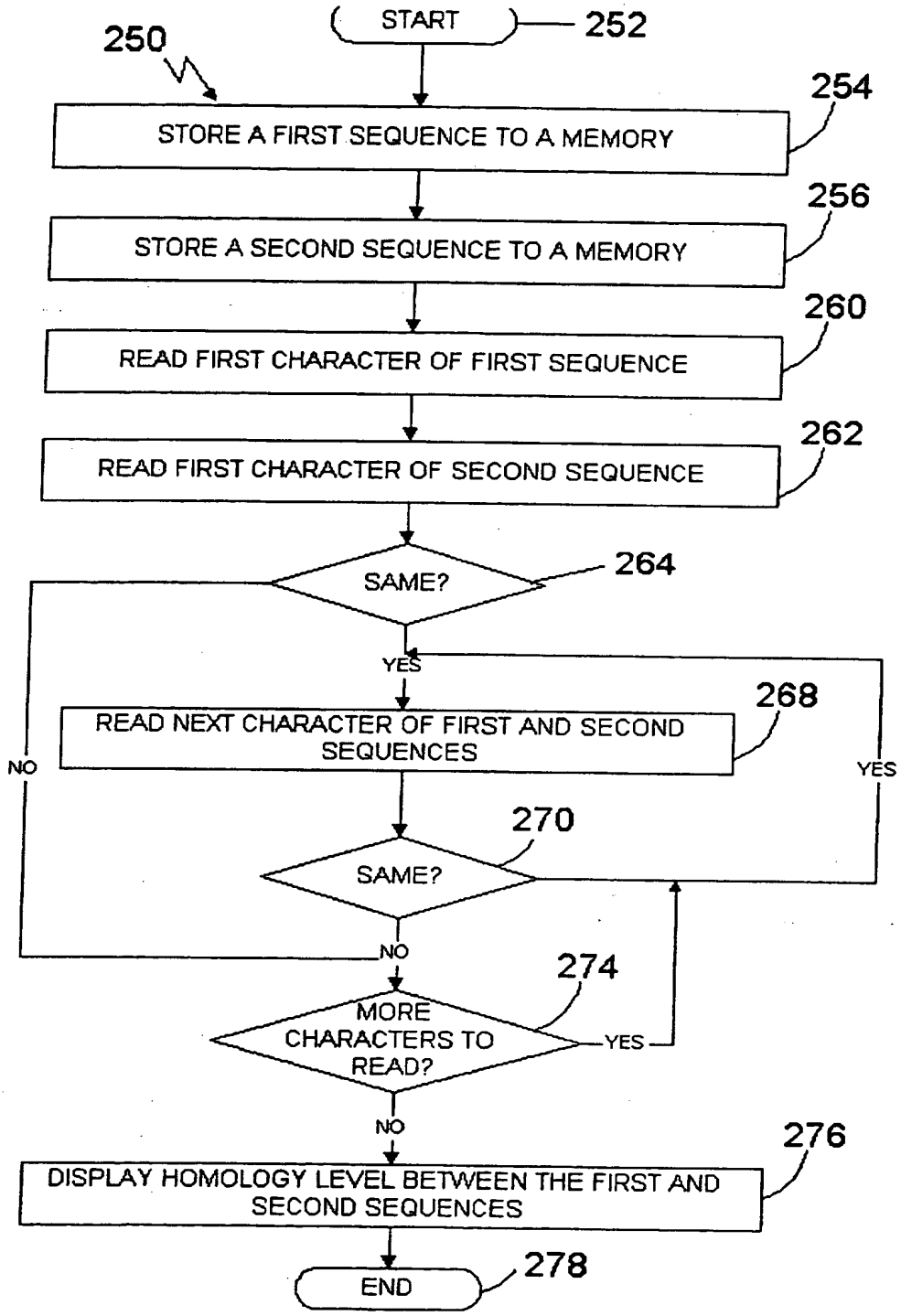

START
252
250
254
STORE A FIRST SEQUENCE TO A MEMORY
256
STORE A SECOND SEQUENCE TO A MEMORY
260
READ FIRST CHARACTER OF FIRST SEQUENCE
262
READ FIRST CHARACTER OF SECOND SEQUENCE
SAME?
264
YES
268
READ NEXT CHARACTER OF FIRST AND SECOND SEQUENCES
NO
YES
270
SAME?
NO
274
MORE CHARACTERS TO READ?
YES
NO
276
DISPLAY HOMOLOGY LEVEL BETWEEN THE FIRST AND SECOND SEQUENCES
END
278

FIG.24

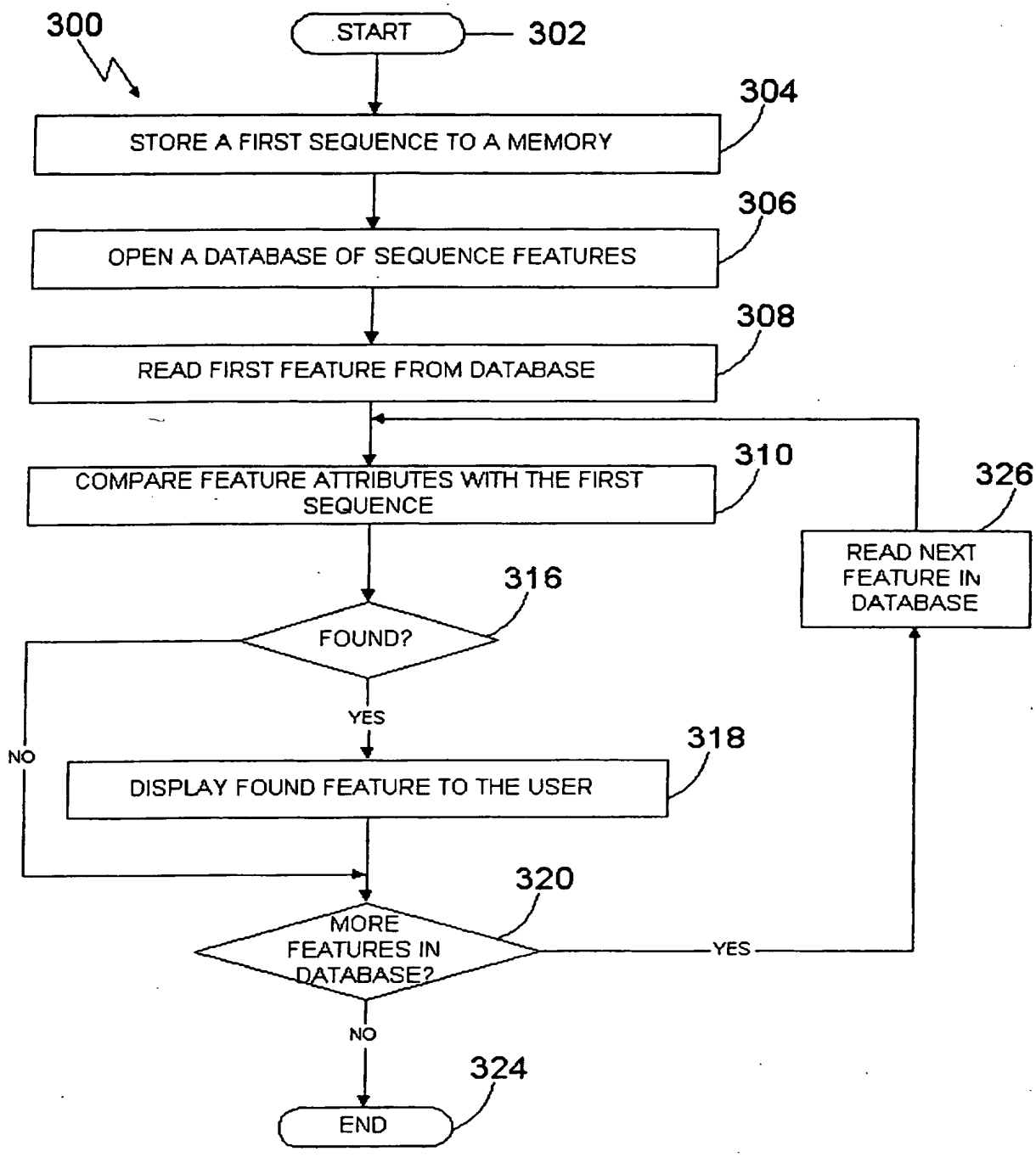

300
START
302
304
STORE A FIRST SEQUENCE TO A MEMORY
306
OPEN A DATABASE OF SEQUENCE FEATURES
308
READ FIRST FEATURE FROM DATABASE
310
COMPARE FEATURE ATTRIBUTES WITH THE FIRST SEQUENCE
326
READ NEXT FEATURE IN DATABASE
316
FOUND?
YES
NO
318
DISPLAY FOUND FEATURE TO THE USER
320
MORE FEATURES IN DATABASE?
YES
NO
324
END

FIG.25

<213> Homo Sapiens

<220>
<221> allele
<222> 501
<223> 12-296-388 : polymorphic base A or G

<220>
<221> misc_binding
<222> 482..500
<223> 12-296-388.mis1

<220>
<221> misc_binding
<222> 502..521
<223> 12-296-388.mis2, potential complement

<220>
<221> primer_bind
<222> 114..134
<223> upstream amplification primer

<220>
<221> primer_bind
<222> 543..563
<223> downstream amplification primer, complement

<220>
<221> misc_binding
<222> 489..513
<223> 12-296-388 potential probe

<220>
<221> misc_feature
<222> 306,378
<223> n=a, g, c or t

<400> 441
catgacacat gtatacatat gtaacaaacc tgcacgttgt gcacatgtat cctagaactt 60
aaagtagaat taaaaaaaaa aaaagaaaat tcactaatta ggtgtgtgac cttgggtaaa 120
gttcaagagc aaagctacac aacttcccta agcctcagtt tcttcatcat taattgggga 180
agtgctctga ggctcaaatg aggtcatgca cataaagcac tcggcatggg cagggcatct 240
gagggaagcc tgaacacttg gcattgggca ttgttattgt tcttcctcag gctcctctca 300
tgttgntctt tctttacagc aaagttggcc cgagatgacc aaattcacat tctcaagcaa 360
caccgacgta aagaactnga aacacggcaa aaacaatatc ggtgagttat gacatcagat 420
cgagtggcca cggggccatg gtttcttcta tctcaagagc atggtatgaa ataaaccctt 480
tccacagtgt atggcctgtc rctgctggaa agtcaccaag gcaccagctc ccagtgtggg 540
tgcacaatcc ccaaacttca ccttacctac aactccagac aagagatgga gaggagagta 600
atgaaaataa ctttggatca accatattgg catgtaggaa gaatgggcag aaaagcttga 660
gtgtacctta atgttctcca ttgtaggggt tgcctcttac ttcttatgtc ttcttataac 720
aattcaatta taaaagtgat acatgatcag caataaaatt gacaactctt cagccaactg 780
accagaggat tcaaaccatt aaaatgagca ataagagaaa atcaatacag aaattaaaat 840
gaataaaaaa ggaaaaatga gactcataaa atgagggaaa atcaatacag aaattaaaat 900
taataaaaaa gaatactatg aacaattgca tgccaacaaa ctagttaaat gaaatgcgca 960
aattcctaga aagacgcaaa ctgccaaaac tgactcatga a 1001

<210> 442
<211> 1001
<212> DNA
<213> Homo Sapiens

<220>
<221> allele

<222> 501
<223> 10-388-379 : polymorphic base C or T

<220>
<221> misc_binding
<222> 481..500
<223> 10-388-379.mis1, potential

<220>
<221> misc_binding
<222> 502..521
<223> 10-388-379.mis2, potential complement

<220>
<221> primer_bind
<222> 123..141
<223> upstream amplification primer

<220>
<221> primer_bind
<222> 542..561
<223> downstream amplification primer, complement

<220>
<221> misc_binding
<222> 489..513
<223> 10-388-379 potential probe

<400> 442
agggccagct caggtatgat gagttggtgc cacagcctga cccagtgatg ggcaggggc 60
agcagttgga gctgtgccca tccagggagc actcgggaga ggagaccaag cagggactct 120
gctcttaggt gaggtcagga gggccatggc cctggctgcc ctctactcag agctcagggt 180
gggcctcgct tttctcctgg tagagcgggt catgaatcac tggcaggaag acctgatgtt 240
tggctaccag ttcctgaatg gctgcaaccc tgtgttgatc cggcgctgca cagagctgcc 300
cgagaagctc ccggtgacca cggagatggt agagtgcagc ctggagcggc agctcagctt 360
ggagcaggag gtccaggtag gggttgatgg gctggggaag tggccaaggt cacagtctgt 420
caggtggaag ccagttcctc ctggccagtg ctcataggcc accaagacgc taactgcagg 480
cccatctggc ctacagcagc ygcttccttt tcctggcagc agtgtcagcc agggtcctgg 540
gcattatgca gactgtcttg tgcaacatca gaggaggaat tgcggggaat gtttctccat 600
gatgctcgag tctgggaaca taatgtcaat attttcacta tcagtatcaa taattacagg 660
agctaccctt gattaggggc ctgtggtggg acaggcattg tgacaggtgc tttacacaca 720
aggtcatcag ttgtcaccca ccttgcaaag gggagaacac tggagaaaga agcagacctt 780
gtgtgagaat aaaaaagggg cacgaggaga aacccaacag aatggttgat tttctcgtta 840
gaaatgctgc tgcccatgct cctgtcgcac cctcctcccc accctcacct gcagactcag 900
ctgcctccgg ctgcaaggct gacctagtct tgagacagaa gaagttcaaa ccaactccac 960
ctggatctgg tggggctcag caacaggcgc tggcccatca g 1001

<210> 443
<211> 1001
<212> DNA
<213> Homo Sapiens

<220>
<221> allele
<222> 501
<223> 10-389-116 : polymorphic base A or G

<220>
<221> misc_binding
<222> 481..500
<223> 10-389-116.mis1, potential

<220>

<222> 481..500
<223> 12-296-119.mis1, potential

<220>
<221> misc_binding
<222> 502..521
<223> 12-296-119.mis2, potential complement

<220>
<221> primer_bind
<222> 383..403
<223> upstream amplification primer

<220>
<221> primer_bind
<222> 812..832
<223> downstream amplification primer, complement

<220>
<221> misc_binding
<222> 489..513
<223> 12-296-119 potential probe

<220>
<221> misc_feature
<222> 575,647
<223> n=a, g, c or t

<400> 450
tatgcagcca taaaaaagga tgaattcatg tcctttgtag ggacatggat gaagctggaa 60
atcacccttc tcagcaaact atcgcaagga caaaaaaaca aacactacat gttctcactc 120
ataggtggga attgaacaat gagaacactt ggacacagga aggggaacat cacacaccaa 180
ggtctgtcat ggggtggggg taagggggag ggatagcatt aggagatata actaatgtaa 240
atgacgagtt aatgggtgca gcacaccaac atgacacatg tatacatatg taacaaacct 300
gcacgttgtg cacatgtatc ctagaactta aagtagaatt aaaaaaaaaa aaagaaaatt 360
cactaattag gtgtgtgacc ttgggtaaag ttcaagagca aagctacaca acttccctaa 420
gcctcagttt cttcatcatt aattggggaa gtgctctgag gctcaaatga ggtcatgcac 480
ataaagcact cggcatgggc rgggcatctg agggaagcct gaacacttgg cattgggcat 540
tgttattgtt cttcctcagg ctcctctcat gttgntcttt ctttacagca aagttggccc 600
gagatgacca aattcacatt ctcaagcaac accgacgtaa agaactngaa acacggcaaa 660
aacaatatcg gtgagttatg acatcagatc gagtggccac ggggccatgg tttcttctat 720
ctcaagagca tggtatgaaa taaacccttt ccacagtgta tggcctgtca ctgctggaaa 780
gtcaccaagg caccagctcc cagtgtgggt gcacaatccc caaacttcac cttacctaca 840
actccagaca agagatggag aggagagtaa tgaaaataac tttggatcaa ccatattggc 900
atgtaggaag aatgggcaga aaagcttgag tgtaccttaa tgttctccat tgtaggggtt 960
gcctcttact tcttatgtct tcttataaca attcaattat a 1001

<210> 451
<211> 1001
<212> DNA
<213> Homo Sapiens

<220>
<221> allele
<222> 501
<223> 12-297-291 : polymorphic base C or T

<220>
<221> misc_binding
<222> 481..500
<223> 12-297-291.mis1, potential

<220>

<221> misc_binding
<222> 502..521
<223> 12-297-291.mis2, potential complement

<220>
<221> primer_bind
<222> 211..229
<223> upstream amplification primer

<220>
<221> primer_bind
<222> 688..707
<223> downstream amplification primer, complement

<220>
<221> misc_binding
<222> 489..513
<223> 12-297-291 potential probe

<400> 451
ctcagtttct atctgtaaag caggataatc actcgctcgg ccattcccag gcccagtgct 60
tcctgagggc ctgttggggg gtgtccactg ttgatagtgc ccaccataat tcagtgtgtc 120
acctgagaat ttaatgtgtc acagcaaata gacaaatata gcatcagccc agtgtacttg 180
gcccagagag gagggaggaa ggtggccatt gtagttccac aaaccttccg aagatatttt 240
gcagtttgac acatgtgacc ctggtatttt cctggaagga ggcctttatc tgtcattctt 300
ccattagtgg cccctgggat gttagaacca cagccatcaa gctttgtctt ttactccctg 360
agttgggtca aggctggagc cctctgcccc ttgttcacct cttatcatat tcaggcttga 420
agccagcacc aggaaagtgg ctgggctggc catagaccag ccagactgca ggcctggccc 480
cctcctgctg tggaagtccc ygtcagatgg cagatggagt ccttgtaaat ctgcctagtg 540
atttcacccc aagtgaaact caggggctct ccaccaggtg cctgcaggtg taggtagggc 600
agggggccag gagtgtccca gtcccctgtg actgcaccac aggtgtgctg gagggtcacc 660
ccactgtcca ctccctcaga gaagctgctc agtctcttcg tgtgggtcgt caccccccagg 720
tagacatcgg gccctccctg tgggacctgc ccagtcctgg aatgcctcct cagcccatct 780
ccctcccctc ttagtccaag gtactgggag cctccctccc atgagagtgt ccctgttagc 840
tcctcccacc ccaccccagc tttcctgtca tctatccccc ggggacacct tcctccctct 900
ccagcccctc agagccatag gaaagctact tcccctgcag cagccacact cagaccaggc 960
ccgggcctct ggtggccgcc cctcagtcct caccgcacac a 1001

<210> 452
<211> 1001
<212> DNA
<213> Homo Sapiens

<220>
<221> allele
<222> 501
<223> 12-298-105 : polymorphic base G or A

<220>
<221> misc_binding
<222> 481..500
<223> 12-298-105.mis1, potential

<220>
<221> misc_binding
<222> 502..521
<223> 12-298-105.mis2, potential complement

<220>
<221> primer_bind
<222> 586..605
<223> upstream amplification primer, complement

This Page Blank (uspto)